White Paper

Grant PR-50205-15

Universal Scripts Project
Deborah Anderson, Project Director
UC Berkeley
22 December 2017

Medefaidrin script, Nigeria

White Paper – Universal Scripts Project – NARRATIVE

1. Project Activities (01/01/2015-09/30/2017)
Summary
The main goal of the Universal Scripts Project (USP) was to oversee the creation of 12 script proposals for scripts not yet in the Unicode Standard, and to shepherd the proposals through the standards committees. This primary goal was accomplished: 13 scripts were written and approved by at least one of the two standards committees. In addition, 9 sets of character additions were approved, and 11 preliminary script proposals were written. Other supporting activities were also completed: submission of script information to the ScriptSource project and Wikipedia, presentations at conferences and meetings, consultation with experts on proposals, and mentoring new proposal authors.

Modifications to Project Activities
The primary changes in the project's activities were modifications to the list of scripts being worked on, due to various circumstances. For example, the Sani Yi and Chuxiong Yi scripts from China were replaced by Khitan Small Script, because experts in China were keen to make progress on the Khitan Small Script.[1]

Also, the archiving of documents was delayed because the eScholarship project at the California Digital Library (CDL) rolled out a new website design in the summer of 2017. The project materials will be uploaded to CDL as part of a new series devoted to script proposals in early 2018.

Matching Funds
The project included a federal matching fund component, which was met through a Google Research Grant and waived indirect costs. The federal matching aspect has been an attractive feature for donors in past NEH grants, but because the Google Grant (and waived IDCs) satisfied the amount to be matched, it was not a great concern for the Project Leader for this grant.

Publicity of Project
The results and ongoing work of the project were publicized through presentations by the Project Director and the project's 2015 post-doc, Anshuman Pandey, at conferences, meetings, and a meetup.[2] In addition, the project was mentioned in articles in *The Atlantic*, the online publishing platform Medium, in the *New York Times* and *The Economist* (online).[3]

2. Accomplishments
The following section lists the specific scripts and sets of characters that were approved, as well as proposals that have still not been approved.

---

[1] The original list included the following 12: Ranjana, Rohingya, Nandinagari, Newa, Siyaq (4), Dogra, Chuxiong Yi, Sani Yi, and Shuishu. As noted, the two Yi scripts were replaced by Khitan Small Script. Shuishu was removed from the list, as China was working on a proposal, and Newa was also removed, as it had been approved in 2014, before the grant started. Additional time will be needed to complete two Siyaq proposals and Ranjana. Because of these changes, the list was modified to include the following new scripts: four historic scripts (Old Sogdian, Sogdian, Makasar and Soyombo) and five modern ones (Medefaidrin, Wancho, Hniaking Puachue Hmong, Gunjala Gondi and Masaram Gondi).

[2] See full list of conference presentations and meetings in Appendix A. The meetup was a San Francisco Globalization meeting in August 2015. Two activities took place on the UC Berkeley campus to help draw attention to the project: a Unicode informational session by the Project Leader and project post-doc Pandey (March 2015), and a poster session at the UC Berkeley Digital Humanities Faire (April 2015).

[3] See Appendix C.

a. Script Proposals – Approved

In the NEH proposal, the projected number of scripts to be approved by at least one standards committee was 12. The project successfully shepherded 13 scripts through the approval process (in the table below, with links to the approved proposals). An approval by a standards committee generally indicates that the proposal is "mature" and provides adequate information about the script and how it works, so it can be implemented in software and fonts.

Wancho [L2/17-067]	Nandinagari [L2/17-162]
Nyiakeng Puachue Hmong [L2/16-070]	Old Sogdian [L2/16-312]
Hanifi Rohingya [L2/16-311]	Medefaidrin [L2/16-101]
Khitan Small Script [L2/16-245]	Indic Siyaq [L2/15-121]
Dogra [L2/15-234]	Gunjala Gondi [L2/15-235]
Masaram Gondi [L2/15-090]	Makasar [L2/15-233]
Soyombo [L2/15-004]	

(Note: Old Sogdian, Hanifi Rohingya, and Nandinagari had funding from NEH and the Unicode Adopt-a-Character program.)

b. Character Additions Approved

The following 9 sets of character additions (to already encoded scripts) were proposed and approved by one of the two standards committees:

Soyombo additions (1 char) [L2/16-016]	Kaithi addition (1 char) [L2/16-097]
Medieval punctuation (5 chars) [L2/16-235, L2/16-220, L2/16-125]	Zanabazar Square additions (3 chars) [L2/15-342, L2/15-341]
Devanagari additions (2 chars) [L2/15-335]	Old Italic letters (3 chars.) [L2/15-181]
Gujarati signs for Arabic (6 chars) [L2/15-103]	N'Ko additions (3 chars) [L2/15-338]
Syriac letters for Garshuni Malayalam (11 chars.) [L2/15-156]	

c. Script Proposals – Preliminary (not yet approved)

The following 11 preliminary proposals were written as part of the project:

Vithkuqi script [L2/17-316]	Palaeohispanic script [L2/17-129]
Proto-Cuneiform [L2/17-157]	Cypro-Minoan [L2/16-179]
Western Cham [L2/16-198]	Mandombe [L2/16-077 and L2/16-019]
Garay [L2/16-069]	Diwani Siyaq [L2/15-066]
Persian Siyaq [L2/15-122]	Eebee Hmong [L2/15-180]
Ottoman Siyaq [L2/15-072][4]	

---

[4] Additional work on the proposal for Ottoman Siyaq was conducted in 2017, with funding from the Unicode Adopt-a-Character program, resulting in a revised proposal (L2/17-348). It was approved by one of the standards committees in October 2017.

d. Character Additions Still in Process (not yet approved)

The following 2 proposals were written as part of the project, but have not yet been approved by the Unicode Technical Committee. More research is required.

- Egyptian Hieroglyphs extensions [L2/17-073, L2/16-257, L2/16-079, andL2/15-240] (related: preliminary proposal for Möller's Egyptian Hieroglyphs [L2/16-250][5])
- Medievalist punctuation [L2/16-125] (Note: Some characters in this omnibus proposal were approved, but others have not yet been approved.)

e. Approved Proposals the Project Leader Assisted On[6]

- Sogdian script [L2/16-371]
- Armenian additions [L2/17-032]

f. Preliminary Proposals the Project Leader Assisted On

- Khwarezmian script [L2/17-054]
- Divehi script [L2/17-292]
- New Hangul additions [L2/17-125]

g. Research Documents on Scripts (not yet approved)

Loma: Updates on encoding the Loma script [L2/17-003, L2/17-059, L2/17-233, and L2/16-201]	Ranjana/Lantsa: Document discussing the encoding for the Ranjana and Lantsa scripts [L2/16-015]
Pau Cin Hau Syllabary: Revised code chart for Pau Cin Hau Syllabary [L2/16-014]	Kulitan: Discussion paper on options for encoding the Kulitan script [L2/15-232]
Garay: Report from the field [L2/17-322]	Bété: Report from the field and preliminary code chart [L2/17-323]

h. Other Documents (including Feedback)

Feedback and other documents for the standards committees included:

Evidence of diaeresis in Armenian [L2/17-315]	Comments on Sandhi marks for Bengali and Newa [L2/17-130]
Feedback on glyphs for Kannada letters Vocalic L and Vocalic LL and their vowel signs [L2/17-127]	Layout and reading order of Mayan Numerals in Mayan historical texts [L2/17-110]
Feedback on Cypro-Minoan [L2/16-265]	Medievalist punctuation characters [L2/16-219]
Gunjala Gondi order [L2/16-200]	Name change for U+16FE0 [L2/16-129]
Consolidated Comments on Al-Dani Quranic Marks [L2/16-102]	Comments on N'Ko Additions [L2/15-287]

---

[5] The following document is a mapping table that is related to the Möller proposal: L2/16-251.
[6] Assistance included providing extensive recommendations and/or reviewing the proposal.

Dissemination
Information on 8 scripts was submitted to [ScriptSource](#), a free online resource on scripts, and 28 updates were made to Wikipedia pages on scripts that were proposed through the project. Seven presentations were made at conferences and meetings (see Appendix A).

Related Work: Unicode Script Ad Hoc
One aspect of the project not mentioned in the original NEH proposal, but one that has been instrumental in helping move forward all script proposals has been the Unicode script ad hoc. The script ad hoc is a small group of Unicode experts who discuss proposals and make recommendations to the standards committees. The meetings are organized and run by the Project Leader and take place at least once per quarter, but often more frequently. The ad hoc reports provide script authors with details on how to improve their proposals and help expedite discussion in the quarterly Unicode Technical Committee meetings. A list of the script ad hoc reports (with links) is contained in the Appendix E.

Meetings with Experts
The project's goal is help get proposals through the standards process, which may involve overcoming any hurdles that may arise. In a number of cases, face-to-face meetings with experts were needed. During the 2015-2017 grant period, the Project Leader organized several meetings with experts to resolve issues and help move proposals forward. These included the following:
- July 2015 meeting in Glasgow, Scotland to discuss with experts whether North Italic was a separate script or if additional characters to Old Italic would cover historic alphabets of northern Italy; the meeting led to a [successful proposal](#) for 3 characters and a [Unicode Technical Note](#)
- October 2015 meeting Tokyo, Japan, to discuss Soyombo and Zanabazar Square scripts questions with Mongolian experts (see [meeting report](#)); the meeting led to the creation of the final script proposals (which were approved by the standards committees)
- July 2016 meeting with Egyptologists and standards representatives in Cambridge, England, to discuss how to progress Egyptian Hieroglyphs (see [report](#))
- August 2016 meeting in Yinchuan, China, to discuss with experts Khitan Small Script and Khitan Large Script (with two reports, see [report #1](#) and [report #2](#)); it led to a successful Khitan Small Script proposal
- November 2016 meeting in Fresno, California, between Project Leader and Cher Vang Kong, who created the Hniaking Puachue Hmong script, to get needed information for a proposal (which was completed and approved in 2017)
- September 2017 meeting in Hohhot, Inner Mongolia, with Mongolian and standards experts to discuss how to overcome current problems in the interchange of Mongolian text and documents (see [report](#))

Meetings with New Script Proposal Authors
One other component of the project was to work with new script proposers, explaining the encoding process and assisting them as needed. The Project Leader arranged meetings between Unicode experts and the following three scholars:
- Carlos Pallán, a graduate student at the University of Bonn who is working on Mayan hieroglyphs; meetings were held at UC Berkeley in April 2015, January 2016 and January 2017 and Mayan hieroglyphs are now part of the 2017-2019 NEH grant

- Noemí Moncunill Martí, a research fellow at the University of Nottingham-Oxford University who is working on the Palaeohispanic scripts; a meeting was held at UC Berkeley in January 2017, and the USP project[7] is now a co-author on the Palaeohispanic proposal
- Christian Casey, a Brown University graduate student; meetings were held in September 2016 at UC Berkeley and April 2017 in Los Angeles, and Casey will be working on a Unicode proposal for Demotic as part of his Ph.D.
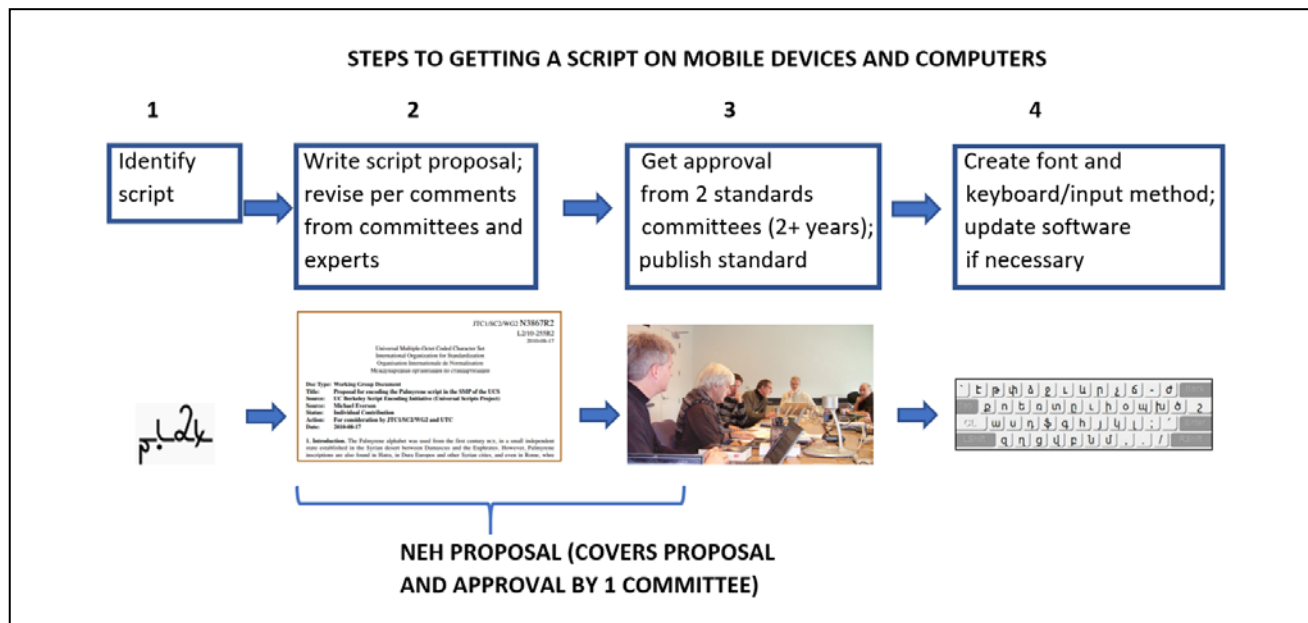
Project Goals Not Achieved and Future Plans
The documents from the project were not uploaded to the eScholarship website during the project timeframe, due to a complete revision of the eScholarship website. The documents and accompanying XML will be uploaded by January 2018.

Many of the script proposals not yet completed have been rolled into a new NEH grant (PR-253360-17), which will run through 2019. Script proposal authoring can be very dependent on outside factors, such as finding experts who respond to email and understand the goals of script encoding, as well as the proposal author finding adequate resource materials. As a result, it is not always easy to predict when a script proposal will be completed and approved. Hopefully many of the script proposals that were worked on as part of the USP project under review here will be completed by the end of 2019.

3. Audiences
The Universal Scripts Project aims to help underserved communities through the development of technical standards for the scripts of historical and modern minority languages. The work done by the Universal Script Project is one critical step in the process of making scripts accessible, but additional steps are needed before the letters and symbols of a script are usable on a computer or mobile device (especially, step 4, below right).

STEPS TO GETTING A SCRIPT ON MOBILE DEVICES AND COMPUTERS

1	2	3	4
Identify script	Write script proposal; revise per comments from committees and experts	Get approval from 2 standards committees (2+ years); publish standard	Create font and keyboard/input method; update software if necessary

NEH PROPOSAL (COVERS PROPOSAL AND APPROVAL BY 1 COMMITTEE)

---

[7] In the Palaeohispanic proposal, as well as in a number of other publications, the project is referred to by the name of the umbrella project, "Script Encoding Initiative" (or "SEI").

The potential audiences for the approved scripts are outlined in the following table. Note that the numbers provided are based on the total number of language speakers, hence it identifies only the "potential" number of users of the new scripts.

SCRIPT	# POTENTIAL USERS	LANGUAGE	GEOGRAPHICAL REGION	NOTES	Time period
MODERN					
Wancho	51,000	Wancho	NE India, Myanmar, Bangladesh		
Hniakeng Puachue Hmong	1000+	White Hmong and Green Hmong	US (Calif.), Laos, etc.	Used by United Christians Liberty Evangelical church members	
Hanifi Rohingya	1.2 million	Rohingya	Myanmar, etc.	Language commonly written with Arabic, Latin or Myanmar scripts	
Medefaidrin	4000	Medefaidrin	Africa (SE Nigeria)	Used by members of Oberi Okaime church	
Gunjala Gondi	300,000	Adilabad Gondi	India		
Masaram Gondi	150,000	Aheri Gondi	India		
HISTORIC					
Dogra	(scholars)	Dogri	India	Devanagari script used today	15C-20C
Indic Siyaq	(scholars)		India and South Asia	Numeric notational system	until mid-20C
Khitan Small Script	(scholars)	Khitan	China and surrounding areas		920-1191
Makasar	(scholars)	Makasar	Indonesia		18C-20C
Nandinagari	(scholars)	Sanskrit, Kannada	India		used 11C-19C
Old Sogdian	(scholars)		Kazakhstan, W China, and surrounding areas		3C-6C
Soyombo	(scholars/liturgical)	Sanskrit, Tibetan, Mongolian	Mongolia and surrounding areas	some modern liturgical use	1686 -

TABLE 1: POTENTIAL AUDIENCES FOR APPROVED SCRIPTS

Working with experts at other institutions
Rather than entire institutions, the project primarily worked with experts within individual institutions (such as linguists, etc.). Table 2 (below) lists experts on the scripts that were approved. Of the 21 listed, 9 are from Asia, 2 from Australia, 2 from Africa, 6 from Europe, and 2 from the U.S. Thirteen are affiliated with universities, 3 with academies of sciences, 1 church, 1 middle school, and 3 other institutions.

The project also has collaborated with the Unicode Consortium. The Unicode Consortium's Adopt-a-Character program funded a meeting on Egyptian hieroglyphs, research on Mayan hieroglyphs, and has paid for work on 3 final script proposals whose preliminary versions were funded through NEH. Unicode also received a Henry Luce Foundation grant, which paid for a portion of three meetings in East Asia related to script proposals.

	CONTACT NAME	INSTITUTIONAL AFFILIATION
Modern Scripts		
Wancho	Stephen Morey	La Trobe University, Melbourne, Australia
	Banwang Losu	Middle school teacher, Arunachal Pradesh, India
Hniakeng Puachue Hmong	Cher Vang Kong	United Christians Liberty Evangelical church, Fresno, CA
Hanifi Rohingya	Muhammad Noor	Rohingya Project, Malaysia
Medefaidrin	Dafydd Gibbon	Universität Bielefeld, Germany
	Moses Ekpenyong and Eno-Abasi Urua	University of Uyo, Nigeria
Gunjala Gondi	Mukund Gokhale	Script Research Institute, Pune
Masaram Gondi	Mark Penny	SIL India
Historic Scripts		
Dogra	Christopher Shackle	SOAS, University of London
Indic Siyaq	Brian Spooner	University of Pennsylvania
Khitan Small Script	Sun Bojun	Chinese Academy of Social Sciences, Beijing
	Wu Yingzhe	Inner Mongolia University, Hohhot, China
	Jing Yongshi	North Nationalities University, China
	Viacheslav Zaytsev	Inst. Oriental Manuscripts, Russian Academy of Sciences, St. Petersburg, Russia
Makasar	Anthony Jukes	La Trobe University, Melbourne, Australia
	Ian Caldwell	University of Leeds
Old Sogdian	Nicholas Sims-Williams	SOAS, University of London
	Yutaka Yoshida	University of Kyoto
Soyombo	Agata Bareja-Starzyńska	University of Warsaw, Poland
	R. Otgonbaatar	Mongolian Academy of Sciences

TABLE 2: LIST OF COOPERATING PARTICIPANTS ON APPROVED SCRIPTS

In addition to the experts listed in Table 2, two others have collaborated with the project[8]:

- Rade Matic, co-owner of the website design studio Drei meiner Kollegen and Assistant Professor at the University of Applied Sciences Darmstadt Faculty of Design, re-designed the project's website in November 2015 and has made improvements to the site *pro bono*
- Thomas Huot-Marchand, Director of the Atelier National de Recherche Typographique (ANRT), Nancy, France, supervised the creation of a font for Palaeohispanic; under Huot-Marchand, ANRT will cooperate on fonts for "missing scripts" in the future, as part of a joint project between USP (/SEI) and ANRT.

---

[8] The Project Leader has also been in contact with Gerry Leonidas who runs the MA program at the University of Reading in Typeface Design. The MA students have created fonts for scripts that were proposed through the USP project. In addition, the Project Leader has continued to collaborate with Johannes Bergerhausen, who runs the decodeunicode project at the University of Applied Sciences, Mainz.

Finally, the project has attracted students to UC Berkeley to work under the Project Leader. Most recently, Zhang Yinghui, a Ph.D. Student from the Graduate School of Chinese Academy of Social Sciences, Beijing, has come for the year (September 2017-September 2018) to learn about encoding historic scripts.

4. Evaluation
The project was evaluated in two different settings: (1) in yearly meetings with the outside evaluator Laura Welcher, and (2) with project board members at annual meetings.

With Laura Welcher, the focus was how to raise the project's profile, locate volunteers and others to help on the project, and how the project can sustain itself in the future. At the annual meetings with board members, the discussion centered on review the list of scripts for the coming year, suggestions to improve the project, and any other issues.

The following highlights comments from the evaluators, with any resulting action by the Project Leader (***in bold italics***):

- The new website, http://linguistics.berkeley.edu/sei/, was found useful, and evaluators had the following suggestions:

    o Make data from the project available on maps and charts (e.g., show a sample of the script with basic information, and add querying capability)
    Result: Basic information on scripts, including geographical info, was collected. Information still needs to be re-proofed and made live. To be done 2018-19.

    o Consider prioritizing scripts on the list of unencoded scripts by giving more weight, for example, to those scripts that have accessible reference materials and users who are available to answer questions.
    Result: Good idea, add to project list of to-dos for 2018-19.

    o Consider sharing data with other projects (such as Langscape project).
    Result: Good idea, worth following up on in 2018-19.

- Having students enter information from the script proposals in Scriptsource website and Wikipedia pages was a good way to disseminate information.
Result: Students did update Wikipedia pages and Scriptsource website entries, which involved review by Project Leader. The Wikipedia pages are not consistent, and more information could be added. Will add to 2018-19 to-do list.

- Another way to promote the project would be to reach out to journalists, and possibly suggest a human-interest story that touches on scripts (and the project).
Result: Project Leader maintained contact with journalist Michael Erard who wrote the NY Times article on script encoding in 2003. This journalist wrote an article in NYT Magazine article that was published Oct. 18, 2017. Journalists from the Economist and The Atlantic contacted the Project Leader (thanks to referrals from a board member and the outside evaluator), and these contacts resulted in two articles. (See Appendix C.)

- To attract new volunteers, consider adding a "help wanted" section of the website. Also, try advertising at UC Berkeley for interns, who could get experience (and perhaps credit).

> ***Result: Good idea, but finding someone with familiarity with the project and script encoding—without requiring too much of the Project Leader's time—has been difficult. Using students for simple tasks has worked out. Nevertheless, it is recognized this could be a good way to recruit new proposal authors or others who can assist on implementing technical features on the website, for example. Add to 2018-19 to-do list.***

- Investigate whether the Open Language Archives Community could include a new field for script code (ISO 15924).
  Result: Laura Welcher and the Project Leader will follow up in 2018-19.

Issues that arose during the project:
- The board was helpful in resolving issues involving personnel involved in the project.
  Result: The Project Leader followed the advice of the board, with success.

Candid review of the project by Project Leader
Strengths
- The project has been extremely successful in the creation of solid proposals that have been vetted by encoding and script experts, and been approved by the standards bodies.
- The impact of the project is significant: once scripts are in the Unicode Standard, the foundation is laid for use of the scripts on computers and mobile devices, for search, and for long-term archiving of texts.
- The established workflow of the project works very well: the proposal is typically written by a seasoned author (or, for a new author, supervised by the Project Leader); the proposal is reviewed by script experts, the Project Leader, and encoding experts in the script ad hoc; the proposal then goes to the standards committees for review and eventual approval.
- The project and Unicode generally have become somewhat more understandable to the general public, and has started to draw interest from journalists.
- Thanks to the success of Unicode and the project, users of unencoded scripts often contact the Project Leader (without prompting or solicitation) and ask for their script to be included in Unicode. (This was not the case when the project originally started in 2002.)
- Collaboration with other projects and institutions, such as the Unicode Consortium, have been a key to the success of the project. Working together with Unicode's Adopt-a-Character program has allowed joint funding of scripts.

Weaknesses
- The success of the project has generated more work than expected. As a result, some technical features, which could aid in making the data more accessible, have not yet been incorporated.
- The number of high-quality proposal authors is very limited. Training new script proposal authors is very time-consuming—but worthwhile, if they work out.
- The scripts that remain to be encoded typically involve extensive research and contact with specialists who are often very hard to reach. Hence, it often takes longer than the usual 2-year timespan of a grant to get a script proposal written, revised, and approved. While the project will continue through another 2-year cycle (thanks to another NEH grant and a large donation), scripts still "in progress" at the end, will still require funding in order for the Project Leader (or her successor) to see scripts through the approval process.
- Fonts and keyboards (or other input methods) are needed in order for the encoded scripts to be used. This aspect has been folded into the 2017-2019 NEH grant, but only for one script block

(Old Italic). Much more work needs to be done. The [Google Noto Fonts](#) projects aims to provide fonts for all the approved scripts, but it is about 5 years behind the approved scripts. Collaboration between the project and ANRT and students in the University of Reading MA program will help. Documentation and/or workshops on how to create fonts and keyboards might be a good option to pursue.

Public response
Discussion with journalists and others from the general public suggests that the goal of the project and Unicode generally, are still somewhat murky to the average person. However, the project's revised website makes the project more accessible. The articles in the press, such as *The Atlantic*, do help to explain how Unicode works in a more understandable way and get most of the facts correct.

5. Continuation of the Project
The project has received another NEH grant (PR-253360-17) for 2017-2019, which will be run essentially in the same manner as the project reviewed in this report. An additional large donation was received in 2016, which will enable the project to go at least through 2020.

The project is run by the Project Leader totally on soft funding. To date, UC Berkeley has made no commitment to continue the project after funding runs out (ca. 2020). At the conclusion of the project at UC Berkeley, the project's documents will be available in the eScholarship section of the California Digital Library. The project website will also be archived as part of an agreement with the UC Berkeley Library. In addition, all proposals and related documents will be archived on the Unicode website.

As noted above, the project has forged a partnership with ANRT (on fonts), decodeunicode (which publicizes information on unencoded scripts), and the Unicode Consortium, which is expected to continue as long as the project lives. Connections with experts at various universities and institutions will also continue during the duration of the project, particularly when questions come up on specific scripts and characters.

6. Long Term Impact
In terms of long-term impact, the project will affect research and teaching by students and scholars at the university who are working with historical and minority language materials. Beyond the classroom, the project will ultimately affect search and archiving of historical and minority language materials on computers and mobile devices worldwide.

The public service aspect of the project has reinforced the view of UC Berkeley as an educational institution dedicated to the improvement of society. At meetings in Asia, the Project Leader's affiliation with UC Berkeley has been viewed very positively by foreign script experts.

The project has brought in a large donation of non-federal support.

7. Grant Products
A list of all the documents produced is contained in Appendix E.

Other publications include the following block introductions by the Project Leader in the Unicode Standard:
- Masaram Gondi, Soyombo, Nushu in [Unicode 10.0](#) (published June 2017)

- Newa, Bhaiksuki, Marchen, Tangut (and Tangut Components) in <u>Unicode 9.0</u> (published July 2016)
- Anatolian Hieroglyphs, Multani, Ahom in <u>Unicode 8.0</u> (published August/September 2015)

Forthcoming products:
- article in January 2018 issue of *Humanities* magazine about the project (and Mayan hieroglyphs).
- forthcoming "World's Writing Systems" poster that specifically mentions the project; the poster is a joint project between this project, decodeunicode and ANRT.

White Paper – Universal Scripts Project – APPENDICES

A. Conference or workshop papers
- Internationalization and Unicode conference November 2016 presentation, "Script Encoding - Part 2: Working with the User Community"
- Digital Humanities 2016, Kraków, Poland, July 2016 (with Carlos Pallán), "Unlocking the Mayan Hieroglyphic Script with Unicode"
- American Research Center in Egypt (ARCE) April 2016 presentation, "Hieroglyphs in Unicode: Exciting Progress Towards a Universal Computing Standard"
- Internationalization and Unicode conference October 2015 panel, "How the Unicode Consortium Works (And How You Can Get Involved)"
- Digital Humanities 2015, Sydney, Australia, July 2015 (with Stephen Morey), "Negotiating the Issues of Encoding and Producing Traditional Scripts on Computers – Working with Unicode"
- American Library Association (ALA) presentation, San Francisco, June 2015, "Character encoding in Unicode, transliteration, and the future of multilingual search" part of a panel on American Library Association panel was entitled "Managing Transliteration of Bibliographic Data"

Meetings attended
- American Oriental Society, Los Angeles, CA, March 2017
- Archaeological Institute of America / Society for Classical Studies joint meeting, San Francisco, January 2016

B. Public acknowledgement of scripts encoded through the project:
http://www.unicode.org/acknowledgements/sei.html

C. Links to articles which mention project (under its umbrella name, "Script Encoding Initiative")
- December 18, 2017: "The great emoji debate" by Glenn Fleishman (*The Economist* online)
  https://www.economist.com/blogs/economist-explains/2017/12/economist-explains-101

- October 18, 2017: "How the Appetite for Emojis Complicates the Effort to Standardize the World's Alphabets" by Michael Erard (*New York Times* magazine)
  https://www.nytimes.com/2017/10/18/magazine/how-the-appetite-for-emojis-complicates-the-effort-to-standardize-the-worlds-alphabets.html?_r=0

- Nov. 16, 2016: "The Alphabet That Will Save a People From Disappearing" by Kaveh Waddell (*The Atlantic*)
  https://www.theatlantic.com/technology/archive/2016/11/the-alphabet-that-will-save-a-people-from-disappearing/506987/

- Sept. 18, 2015: "Unicode: A story of corruption, connection, and smiling poo" by Maggie Shafer (Medium publishing platform)
  https://medium.com/@maggieshafer/unicode-a-story-of-corruption-connection-and-smiling-poo-598295e4af9d

D. Representative samples of completed work
The proposal for the Makasar script is appended (pp. 19-50).

E. List of Publications
a. Script Ad Hoc Reports

2017

L2/17-288 Ad Hoc Recommendations on PRI 352 and PRI 353 (Anderson et al.) 2017-08-03

L2/17-255 Recommendations to UTC #152 July-August 2017 on Script Proposals (Anderson et al.) 2017-07-29

L2/17-153 Recommendations to UTC #151 May 2017 on Script Proposals (Anderson et al.) 2017-05-07

L2/17-037 Recommendations to UTC #150 January 2017 on Script Proposals (Anderson et al.) 2017-01-23

2016

L2/16-342 Recommendations to UTC #149 November 2016 on Script Proposals (Anderson et al.) 2016-11-07

L2/16-266 Comments on Mongolian, Small Khitan, and other WG2 #65 documents (Deborah Anderson, et al.) 2016-09-26

L2/16-216 Recommendations to UTC #148 August 2016 on Script Proposals (Anderson et al.) 2016-08-01

L2/16-156 Recommendations to UTC #147 May 2016 on Script Proposals (Anderson et al.) 2016-05-09

L2/16-155 Script Ad Hoc Recommendations on Tamil Fractions and Symbols (and Spreadsheet) (Anderson et al.) 2016-05-09

L2/16-037 Recommendations to UTC #146 January 2016 on Script Proposals 2016-01-24

2015

L2/15-312 Recommendations to UTC #145 November 2015 on Script Proposals (Anderson et al.) 2015-11-03

L2/15-204 Recommendations to UTC #144 July 2015 on Script Proposals (Anderson et al.) 2015-07-27

L2/15-149 Recommendations to UTC #143 May 2015 on Script Proposals (Anderson et al.) 2015-05-04

L2/15-045 Recommendations to UTC #142 February 2015 on Script Proposals (Anderson et al.) 2015-02-01

b. Proposals, Reports, Feedback and Comments Documents
Note: * indicates work was supported by the Unicode Adopt-a-Character program, with assistance from the Project Leader

2017

L2/17-347 Mongolian Ad Hoc Report (Hohhot, Inner Mongolia) (WG2 N4893) (Anderson et al.) 2017-09-28

L2/17-323 Report on the Bété script 2017 (WG2 N4876) (Riley) 2017-09-12

L2/17-322 Report on the Garay script 2017 (WG2 N4875) (Riley) 2017-09-12

L2/17-315 Evidence of diaeresis in Armenian (Deborah Anderson and Yury Golev) 2017-09-11

L2/17-316 Preliminary proposal for encoding the Vithkuqi script (WG2 N4854) SEI / Michael Everson 2017-09-11

L2/17-254 SEI Liaison Report (Anderson) 2017-07-29

* L2/17-292 Proposal to encode Divehi (Pandey) 2017-09-29

*L2/17-235 Proposal to encode JIHVAMULIYA and UPADHMANIYA for Soyombo (Pandey) 2017-07-24

L2/17-233 Cumulative chart of the Loma script (WG2 N4837) (SEI/Everson) 2017-07-24

*L2/17-162 Final proposal to encode Nandinagari in Unicode (Pandey) 2017-05-09

L2/17-157 Proposal to encode Proto-Cuneiform in the SMP (WG2 N4797) (SEI/Everson) 2017-05-09

L2/17-154 SEI Liaison Report (Anderson) 2017-05-07

L2/17-130 Comments on L2/16-322 and L2/16-383, Sandhi marks for Bengali and Newa (Anderson) 2017-05-01

L2/17-129 Proposal to encode the Palaeohispanic script (Joan Ferrer, et al) 2017-05-01

L2/17-127 Feedback and Response on L2/17-041: Request to change glyphs of Kannada letters Vocalic L and Vocalic LL and their vowel signs (Anderson) 2017-05-01

L2/17-125 Preliminary Proposal for Encoding New Hangul (S. and S. Oh) 2017-05-01

L2/17-110 Mayan Numerals: Layout and reading order in Mayan historical texts (Pallan and Anderson) 2017-04-24

L2/17-073 New draft for the encoding of an extended Egyptian Hieroglyphs repertoire (Hieroglyphica based) (17MB, and associated database PDF snapshot) (Suignard) 2017-03-23

L2/17-067 Proposal to encode the Wancho script (revised; WG2 N4787R) (Everson) 2017-10-23

L2/17-059 Towards an encoding of the Loma script in the SMP (Everson) 2017-02-28

L2/17-055 Preliminary proposal to encode the Elymaic script (Pandey) 2017-02-03

*L2/17-054 Proposal to encode the Khwarezmian script (revised) (Pandey) 2017-07-24

 L2/17-042 Preliminary proposal to encode the Wancho script (WG2 N4785) 2017-01-22

L2/17-039 SEI Liaison Report (Anderson) 2017-01-22

L2/17-032 Armenian Phonetic Characters in Unicode (Baronian) 2017-01-19

L2/17-003 Further updates on encoding the Loma script in the SMP (WG2 N4779) (SEI/Everson) 2017-01-17

L2/17-002 Proposal to encode the Nyiakeng Puachue Hmong (SEI/Everson) 2017-01-17

2016

*L2/16-371 Revised proposal to encode the Sogdian script (revised) (Pandey) 2017-01-25

L2/16-341 SEI Liaison Report (Anderson) 2016-11-06

* L2/16-312 Proposal to encode the Old Sogdian script (revised) (Pandey) 2017-01-25

* L2/16-311 Revised proposal to encode Hanifi Rohingya (revised) (Pandey) 2017-01-25

* L2/16-310 Revised proposal to encode Nandinagari (Pandey) 2016-11-29

L2/16-297 Mongolian ad hoc report (WG2 N4769) (Anderson) 2016-10-27

L2/16-267 Preliminary proposal to encode Proto-Cuneiform in the SMP (WG2 N4760) (SEI, Michael Everson, Laura F. Hawkins) 2016-09-26

L2/16-265 Feedback on Cypro-Minoan (based on Script Ad Hoc comments in L2/16-216) (Anderson) 2016-09-26

L2/16-257 Source analysis of an extended Egyptian Hieroglyphs repertoire (Hieroglyphica) (Suignard) 2016-09-20

L2/16-251 Preliminary Mapping table of Möller's Egyptian Hieroglyphs (WG2 N4742 ; see also associated spreadsheet) (Anderson) 2016-09-12

L2/16-250 Preliminary proposal to encode Möller's Egyptian Hieroglyphs in the SMP (WG2 N4741) (Everson) 2016-09-14

L2/16-245 Final proposal to encode the Small Khitan Script in the SMP (revised; WG2 N4738r2) (SEI / Michael Everson, et al) 2016-09-26

L2/16-244 Summary of Meeting on Khitan Scripts, 22 August 2016 (Yinchuan, China) - Ad Hoc Report #2 (WG2 N4737) (Anderson et al.) 2016-09-06

L2/16-243 Summary of Meeting on Khitan Scripts, 20 August 2016 (Yinchuan, China) - Ad Hoc Report #1 (WG2 N4736) (Anderson et al.) 2016-09-06

L2/16-235 Proposal to encode PARAGRAPHUS and PUNCTUS ELEVATUS (Anderson and Everson) 2016-08-05

L2/16-220 Proposal for Medieval Comma (Anderson/Everson) 2016-08-01

L2/16-219 Feedback and Evidence on Medievalist punctuation characters (Anderson) 2016-08-01

L2/16-218 Brief Report from Cambridge meeting of Egyptologists and Update (Anderson) 2016-08-01

L2/16-217 SEI Liaison Report (Anderson) 2016-08-01

L2/16-201 Update on encoding the Loma script in the SMP (SEI / Michael Everson) 2016-08-01

L2/16-200 Gunjala Gondi order (Anderson) 2016-07-21

L2/16-198 Proposal to encode Western Cham in the SMP (WG2 N4734) (SEI / Everson) 2016-08-01

L2/16-179 Revised proposal to encode the Cypro-Minoan script in the SMP (WG2 N4733) (SEI / Everson) 2016-07-22

L2/16-154 Script Encoding Initiative Liaison report (Anderson) 2016-05-09

L2/16-129 Name change for U+16FE0 (Anderson and Whistler) 2016-05-05

L2/16-125 Revised Proposal to add Medievalist punctuation characters (WG2 N4726) (SEI / Everson) 2016-05-05

L2/16-102 Consolidated Comments by Mansour, Evans, and Abudena on Al-Dani Quranic Marks (Anderson) 2016-05-02

L2/16-101 Proposal for encoding the Medefaidrin (Oberi Okaime) script in the SMP (replaces L2/16-020) (Rovenchak, et al.) 2016-05-02

L2/16-097 Proposal to encode Kaithi Number Sign Above (revised) (SEI / Yang)   2016-05-11

L2/16-089 Proposal to encode the Cypro-Minoan script in the SMP (WG2 N4715) (Everson) 2016-05-02

L2/16-082 Proposal to encode seven Cham characters in the BMP (WG2 N4711) (Everson) 2016-04-25

L2/16-079 Preliminary draft for the encoding of an extended Egyptian Hieroglyphs repertoire (Suignard) 2016-04-20

L2/16-077 Proposal for encoding the Mandombe script (2nd rev, replaces L2/15-118) (Rovenchak, et al.) 2016-07-19

L2/16-070 Preliminary proposal for encoding the Cher Vang Hmong script in the SMP (Everson) 2016-03-29

L2/16-069 Proposal for encoding the Garay script in the SMP (Everson) 2016-03-29

L2/16-033 Script Encoding Initiative Liaison report (Anderson) 2016-01-22

L2/16-028 Preliminary draft of the extended Egyptian Hieroglyphs repertoire (Suignard) 2016-01-21

L2/16-020 Proposal for encoding the Medefaidrin (Oberi Okaime) script in the SMP (Rovenchak) 2016-01-19

L2/16-019 Proposal for encoding the Mandombe script in the SMP (Rovenchak) 2016-01-19

L2/16-017 Proposal to Encode Arabic Siyaq Numbers in Unicode (Pandey) 2016-01-19

L2/16-016 Proposal to encode the Soyombo mark PLUTA (Pandey) 2016-01-19

L2/16-015 Towards an encoding for the Ranjana and Lantsa scripts (Pandey) 2016-01-19

L2/16-014 Revised code chart for the Pau Cin Hau Syllabary (Pandey) 2016-01-19

L2/16-002 Proposal to encode the Nandinagari script (Pandey) 2016-01-07

2015

L2/15-342 Proposal to Encode a Subjoiner for Zanabazar Square (Pandey) 2015-12-29

L2/15-341 Proposal to encode additional head marks for Zanabazar Square (Pandey) 2015-12-29

L2/15-340 Unification of 'Diwani' and 'Ottoman' Siyaq Numbers (Pandey) 2016-01-07

L2/15-338 Proposal to encode four N'Ko characters in the BMP (Everson) 2016-01-07

L2/15-337 Proposal to Encode the Zanabazar Square Script (replaces L2/14-024) (Pandey) 2015-12-31

L2/15-336 Comments on L2/15-249 regarding Soyombo and Zanabazar Square punctuation (Pandey) 2015-12-10

L2/15-335 Proposal to encode the Devanagari letter and vowel sign AY (Pandey) 2015-12-10

L2/15-331 Proposal to encode JIHVAMULIYA and UPADHMANIYA for Soyombo (Pandey) 2015-12-07

L2/15-327 Proposal to add Medievalist punctuation characters (revised) (Everson) 2016-01-25

L2/15-298 Proposal for encoding the Medefaidrin (Oberi Okaime) script in the SMP (replaces L2/15-117R2) (Rovenchak, et al.) 2015-11-01

L2/15-297 Proposal for encoding the Mandombe script in the SMP (replaces L2/15-118R) (Rovenchak, et al.) 2015-11-01

L2/15-289 Script Encoding Initiative Liaison report (Anderson) 2015-10-31

L2/15-287 Comments on N'Ko Additions (Anderson and Everson) 2015-10-31

L2/15-283 Nushu Ad Hoc Report (WG2 N4693) (Anderson) 2015-10-31

L2/15-278 Proposal to encode the Hanifi Rohingya script in Unicode (revised) (Pandey) 2016-01-07

L2/15-249 Summary of Ad Hoc Meeting on Two Historical Scripts from Mongolia (Tokyo, Japan) (revised) (Anderson) 2015-12-23

L2/15-247 Revised code chart and names list for Soyombo (Pandey) 2015-10-20

L2/15-246 Revised code chart and names list for Zanabazar Square (Pandey) 2015-10-20

L2/15-244 Script Encoding Initiative Liaison report (to WG2) (Anderson) 2015-10-16

L2/15-240 Preliminary draft of the Ptolemaic repertoire (A: Man and his occupations) (Suignard) 2015-10-12

L2/15-235 Proposal to encode the Gunjala Gondi script (Pandey) 2015-11-03

L2/15-234 Proposal to encode the Dogra script (supersedes L2/15-213; revised) (Pandey) 2015-11-05

L2/15-233 Proposal to encode the Makasar script (supersedes L2/15-179) (Pandey) 2015-10-07

L2/15-232 Towards an encoding for Kulitan (Pandey) 2015-10-07

L2/15-213 Preliminary proposal to encode the Dogra script (Pandey) 2015-07-30

L2/15-205 Script Encoding Initiative Liaison report (Anderson) 2015-07-27

L2/15-181 Proposal for the addition of three Old Italic characters (revised; WG2 N4669R) (Everson) 2015-07-28

L2/15-180 Preliminary proposal to encode the Eebee Hmong script (Everson) 2015-07-27

L2/15-179 Proposal to Encode the Old Makassarese Script (supercedes L2/15-100) (Pandey) 2015-07-20

L2/15-156 Proposal to Encode Syriac Letters for Garshuni Malayalam (Pandey) 2015-06-08

L2/15-146 Script Encoding Initiative Liaison report (Anderson) 2015-05-01

L2/15-122 Proposal to Encode Persian Siyaq Numbers (revised) (Pandey) 2016-01-07

L2/15-121 Proposal to Encode Indic Siyaq Numbers (revised) (Pandey) 2015-10-08

L2/15-118 Preliminary proposal for encoding the Mandombe script (revised) (Rovenchak) 2015-07-20

L2/15-117 Preliminary proposal for encoding the Medefaidrin (Oberi Okaime) script (revised) (Rovenchak) 2015-07-29

L2/15-103 Revised Proposal to Encode Gujarati Signs for the Transliteration of Arabic (replaces L2/14-131) (Pandey) 2015-04-02

L2/15-100 Preliminary Proposal to Encode the Makassarese Bird Script (Pandey) 2015-06-25

L2/15-094 Response to Mongolian and Japanese comments on Soyombo and Zanabazar Square (Pandey) 2015-03-12

L2/15-090 Proposal to Encode the Masaram Gondi Script (replaces L2/15-005 ; revised) (Pandey) 2015-06-08

L2/15-089 Preliminary Proposal to Encode the Old Sogdian Script (Pandey) 2015-11-03

L2/15-088 Preliminary Proposal to Encode Syriac Letters for Garshuni Malayalam (Pandey) 2015-04-21

L2/15-086 Preliminary Proposal to Encode the Gunjala Gondi Script (Pandey) 2015-02-20

L2/15-074 Proposal to Encode the SIYAQ NUMBER MARK for Arabic (Pandey) 2015-02-03

L2/15-072 Proposal to Encode Ottoman Siyaq Numbers (revision 2) (Pandey) 2015-12-22

L2/15-066 Proposal to Encode Diwani Siyaq Numbers (revised) (Pandey) 2015-12-10

L2/15-047 CJK Slavonic transcription characters (Anderson) 2015-02-01

L2/15-046 Script Encoding Initiative Liaison report (Anderson) 2015-02-01

L2/15-005 Proposal to Encode the Gondi Script (Pandey) 2015-01-27

L2/15-004 Proposal to Encode the Soyombo Script (WG2 N4655 ; revised) (Pandey) 2015-12-07

Proposal to encode the Makasar script in Unicode

Anshuman Pandey
Department of Linguistics
University of California, Berkeley
Berkeley, California, U.S.A.
anshuman.pandey@berkeley.edu

November 2, 2015

1   Introduction

This is a proposal to encode the 'Makasar' script in the Unicode standard. The script was described by Christopher Miller in Unicode Technical Note #35 "Indonesian and Philippine Scripts and Extensions", where it was identified as "Makassarese *jangang-jangang* bird script" and recommended for encoding in the standard (2011: 43–46, 51). A preliminary proposal to encode the script under the name "Makassarese Bird Script" was presented earlier (L2/15-100). That was superseded by another proposal (L2/15-179), in which the designation for the script was changed to 'Old Makassarese'. The present proposal is a revision of L2/15-179 and incorporates feedback from experts and suggestions made in "Recommendations to UTC #144 July 2015 on Script Proposals" (L2/15-204). The major differences from earlier proposals are:

- Change of the script identifier to 'Makasar' (see section 3.2).
- Reallocation of the block within the SMP from U+11880..1189F to U+11ED0..11EFF.
- Enlargement of the block to accommodate the potential encoding of digits in the future.
- Expansion of the discussion on digits (see section 4.5).
- Removal of the end-of-text character (*tammat*) from the proposed repertoire (see section 4.7).
- Improvements to glyphs for the proposed characters.

2   Background

The Makasar script was used historically in South Sulawesi, Indonesia for writing *basa mangkasara'* or Makasar (ISO 639-3: mak), a Malayo-Polynesian language currently spoken by 2.1 million people. The script was maintained for official purposes in the kingdoms of Makasar in the 17th century. It was used for writing a number of historical accounts, such as the "Chronicles of Gowa and Tallo'". Metal types for the script were developed in the 19th century (see figures 5 and 7).

The script is known indigenously in the Makasar language as *ukiri' jangang-jangang* or *huruf jangang-jangang* "bird letters" and in Bugis as *uki' manu'-manu'*, which also means "bird letters". It is referred to as "Old Makassarese" and "Makassarese bird script" in scholarly works written in English (ie. Jukes 2014: 1; Cummings 2007: 13). However, there is consensus among experts that 'Makasar' is the most acceptable name for the script. The origins of the name *jangang-jangang* are unclear. Birds are considered carriers of

Figure 1: The location of South Sulawesi province in Indonesia. Source: Wikimedia Commons (https://commons.wikimedia.org/wiki/File:South_Sulawesi_in_Indonesia.svg)

communication in the traditions of South Sulawesi (Rahman 2014). The idea of 'bird letters' may, therefore, be a symbolic extension of that concept to the practice of writing as a vehicle of communication. The name may also refer to the shapes of letters, several of which depict silhouettes of birds in various poses. However, it is unclear if the letter forms were deliberately designed to resemble birds, or if the writing of letters in such shapes was influenced by the name of the script.

The Makasar script is one of two Indic writing systems used for representing the Makasar language. The other is *lontara' beru* "new writing", which is known commonly as 'Bugis' or 'Buginese' (see figure 12). The Buginese script is also referred to as the Bugis-Makassar script because of its usage for writing both the Bugis and Makasar languages. The character repertoire of the Makasar script is similar to that of Buginese. The key difference is that Makasar lacks letters that correspond to the Buginese pre-nasalized clusters /ŋka/, /ɲca/, /mpa/, /nra/ and the consonant /h/. A shared feature of the two, or rather deficiency, is that syllable codas are not represented in either script. A comparison of the two scripts is given in figures 1–3. A folio showing usage of the two scripts in a single source is given in figure 4.

The closest relative of Makasar appears to be the Rejang script used in South Sumatra (see figure 11). There are some generic resemblances between the two in terms of letter forms and character repertoire. The nature of their connection, however, is not yet completely known. Makasar also bears some resemblance to Buginese in terms of letter forms and repertoire, but there is no evolutionary relationship between the two.

The script ceased to be used by the 19th century. By this time the Buginese script has begun to become used for writing the Makasar language. At present, there are no native users of the script. According to Anthony Jukes, "there are now no Makassarese [speakers] who can read it [...] even those well versed in reading *lontara'* in Bugis [Buginese] script, need to have old Makassarese *lontara'* transliterated for them before attempting to interpret them" (2014: 6).

In his final proposal for encoding Buginese in Unicode (L2/03-191), Michael Everson made reference to an "older alphabet" for the Makasar language that was described by the Dutch scholar B. F. Matthes (1858). Everson characterized this script as having "different shapes for the letters", but concluded that it could be considered a stylistic variant of Buginese, as "the difference seems to be a change in font style only" (Everson 2003: 1). Although Matthes only refers to the script as "oude schrift" and does not provide a distinct name for it (1858: 12), it is clear from a short example that this 'old script' is in fact the Makasar script. As the present proposal illustrates, there is sufficient justification to encode the Makasar script separately in Unicode, particularly on account of its distinctive letterforms, attestation in historical sources, and concurrent usage with the Buginese script.

2

3   Script Details

3.1   Structure

The Makasar script is an alphasyllabary that is written from left to right. It is based upon the Brahmi model and is related to various scripts of Indonesia and Philippines. The only independent vowel letter is MAKASAR LETTER A, which has the default value /a/, but also functions as a vowel carrier. Vowels are represented using dependent combining signs. These signs are written with the vowel carrier for expressing independent forms of vowels. Each consonant possesses the inherent vowel /a/. The inherent vowel is changed by applying a vowel sign to a consonant. There is no means for silencing the inherent vowel, such as the VIRAMA of various Indic scripts. Vowel signs may occur to the left, right, above, and below a consonant letter. Two vowel signs may occur with a base letter.

The script has a system for abbreviating syllables and reduplicating onset consonants. Abbreviation of syllables is performed by doubling the vowel sign of a base consonant (see section 4.8). Reduplication of an onset consonant is marked using a placeholder, which also functions as a vowel carrier (see 4.4).

The structure of a Makasar orthographic syllable is:

$$\mathbf{V_{carrier}} \mid \mathbf{C} \mid \mathbf{C_{placeholder}} \; [\mathrm{V_{sign\text{-}1}}] \; [\mathrm{V_{sign\text{-}2}}]$$

Various forms of punctuation are used (see section 4.6). Words are generally separated using spaces. Sentences are delimited using three vertical dots, text sections are marked using a triangle consisting of six dots, and end of text is often marked using a stylistic rendering of the Arabic word *tammat* 'it is complete'.

3.2   Script name

There is no standard or conventional English name for the script. Based upon discussions with experts, 'Makasar' has been assigned as the Unicode identifier for the script block. The designation 'Makasar' is the modern, preferred form of the English terms 'Makassar', 'Makassarese', 'Macassar', 'Macassarese', 'Mangkasar'. Of these other spellings 'Makassarese' is the most commonly known and, in fact, aligns morphologically with the identifiers assigned to other Indonesian scripts in Unicode: Balinese, Buginese, Javanese, Sundanese. However, the form 'Makasar' is reflective of indigenous and scholarly preferences, and is also used as the formal name of the Makasar language.

For this reason, 'Makasar' is the most suitable identifer for the script in Unicode. It is more descriptive than the indigenous names '*ukiri' jangang-jangang*' and '*huruf jangang-jangang*' within an English context and it facilitates easier identification of the script within a global environment. It is also more suitable than "Makassarese bird script", which is a translation of *jangang-jangang* that does not adequately capture the semantic nuances of the indigenous name. Furthermore, 'Makasar' is preferred over 'Old Makasar' on grounds of accuracy: the descriptor 'old' suggests that there is a 'new' form of the script; such a form does not exist and the script is not used at present.

The indigenous name '*ukiri' jangang-jangang*' and the alternate English names 'Old Makassarese' and 'Makassarese Bird Script' have been specified as aliases for the script block in the names list.

3.3   Character repertoire

The character repertore for Makasar consists of 18 consonant letters, 4 combining vowel signs, 1 consonant reduplicator, and 2 punctuation marks. Representative glyphs for the proposed characters are based

upon forms used in manuscripts. Character names are patterned upon those used for Buginese characters in Unicode. The ordering of letters also follows that of the Buginese block.

4 Proposed Encoding

4.1 Encoding model

The chief complexity of the Makasar script is the visual ordering of the ꡝ VOWEL SIGN E. Although the vowel represented by this sign is pronounced after a consonant, the sign is written before the consonant. This prepending behavior is identical to that of the corresponding character in Buginese, ꡝ U+1A19 BUGINESE VOWEL SIGN E. There are two possible models for managing such behavior:

- *Logical order*   This approach follows the current model for Buginese. The VOWEL SIGN E would be encoded as a combining sign. In an encoded sequence it would be placed in its logical position after a base consonant, but it would be prepended to the base consonant in the visual output. The rendering engine would handle the repositioning of the sign.

  <KA, VOWEL SIGN E>   →

  <KA, VOWEL SIGN E, KA, VOWEL SIGN E>   →

  Placing the vowel sign manually before the consonant would result in incorrect rendering:

  <VOWEL SIGN E, KA>   →

  <VOWEL SIGN E, KA, VOWEL SIGN E, KA>   →

- *Visual order*   This approach requires manual placement of the VOWEL SIGN E before the consonant in the encoded sequence. There would be no need for rendering. Accordingly, the sign would be encoded as a regular letter or mark, because combining signs cannot occur before the base letter to which they attach; they must follow the base. In this model the vowel mark would be used as follows:

  <VOWEL MARK E, KA>   →

  <VOWEL MARK E, KA, VOWEL MARK E, KA>   →

  Placing this vowel mark after the consonant letter would result in incorrect rendering:

  <KA, VOWEL MARK E>   →

  <KA, VOWEL MARK E, KA, VOWEL MARK E>   →

Of the above, the logical model is considered the more advantageous and is adopted here. It enables the VOWEL SIGN E to be treated properly as a combining sign like the other vowel signs of the script, instead of as a letter or mark. This model also provides for easier identification of syllables, searching, and collation. Additionally, the encoding for Buginese in Unicode is based upon the logical model. Given the relationship between the two scripts and the potential overlap of their user communities, it is practical that the model for the Makasar script be the same as that for Buginese.

4.2 Consonants

Eighteen consonant letters are proposed for encoding:

	Character name	Phonetic value
ᨀ	MAKASAR LETTER KA	/k/
ᨁ	MAKASAR LETTER GA	/g/
ᨂ	MAKASAR LETTER NGA	/ŋ/
ᨄ	MAKASAR LETTER PA	/p/
ᨅ	MAKASAR LETTER BA	/b/
ᨆ	MAKASAR LETTER MA	/m/
ᨈ	MAKASAR LETTER TA	/t/
ᨉ	MAKASAR LETTER DA	/d/
ᨊ	MAKASAR LETTER NA	/n/
ᨌ	MAKASAR LETTER CA	/t͡ʃ/
ᨍ	MAKASAR LETTER JA	/d͡ʒ/
ᨎ	MAKASAR LETTER NYA	/n/
ᨐ	MAKASAR LETTER YA	/j/
ᨑ	MAKASAR LETTER RA	/r/
ᨒ	MAKASAR LETTER LA	/l/
ᨓ	MAKASAR LETTER VA	/w/
ᨔ	MAKASAR LETTER SA	/s/
ᨕ	MAKASAR LETTER A	/a/, 0

The ᨕ MAKASAR LETTER A is a vowel carrier. When a combining vowel sign is attached to it, it adopts the phonetic value of the sign and represents an independent vowel.

The ᨓ MAKASAR LETTER VA is named as 'VA' despite the fact that the consonant is pronounced as /w/ in the Makasar language. The name for the character was selected to align with the name for ᨓ U+1A13 BUGINESE LETTER VA.

Variant forms of consonants are attested in several manuscript sources (see figure 2). These are glyphic variants and do not require separate encoding. The characters with the most distinctive variants are shown below:

	Regular	Variant
KA	ᜊ	ᜉ
DA	ᜇ	ᜑ
RA	ᜍ	ᜒ
YA	ᜌ	ᜋ
SA	ᜐ	᜔
A	ᜀ	᜕

4.3 Vowel signs

Four combining vowel signs are proposed for encoding:

	Character name	Phonetic value
◌̇	MAKASAR VOWEL SIGN I	/i/
◌̣	MAKASAR VOWEL SIGN U	/u/
꠆◌	MAKASAR VOWEL SIGN E	/e/
◌꠆	MAKASAR VOWEL SIGN O	/o/

Independent forms of vowels are represented by attaching vowels signs to ᜀ MAKASAR LETTER A, as shown below. These vowel signs combine with other consonant letters in the same way to form syllables.

a	ᜀ	<ᜀ A>
i	ᜀ̇	<ᜀ A, ◌̇ VOWEL SIGN I>
u	ᜀ̣	<ᜀ A, ◌̣ VOWEL SIGN U>
e	꠆ᜀ	<ᜀ A, ꠆◌ VOWEL SIGN E>
o	ᜀ꠆	<ᜀ A, ◌꠆ VOWEL SIGN O>

In adherence to the recommended encoding model, the ꠆◌ VOWEL SIGN E is placed after the base consonant in the encoded sequence, but reordered before the consonant in the visual output by the rendering engine.

Up to two vowel signs may combine with a base letter, particularly for the abbreviation of syllables as discussed in section 4.8.

6

4.4 Consonant reduplicator

The ᝊ MAKASAR ANGKA is used for reduplicating the onset consonant of the previous syllable in a word (see also the description in figure 5). Its usage is based upon a convention opposite that of the doubling of vowel signs for the abbreviation of syllables (see section 4.8). As there is no sign or other means for marking the inherent vowel of a consonant, it is not possible to abbreviate two contiguous syllables consisting of identical consonants by doubling their vowel signs. This applies solely to cases where the onset consonant and the consonant of the following syllable are identical. In such a case, the consonant following the onset is replaced with the ANGKA.

The usage of ANGKA is illustrated in the following examples. The boxed text in the excerpt below is ᝊᝈ, which is the syllable ᝈ <ᝈ RA, ◌ VOWEL SIGN U> followed by ᝊ ANGKA:[1]
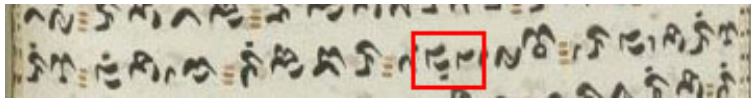
This text is to be read as ᝈᝈ *rura*. As shown, the ANGKA reduplicates the onset consonant RA of the previous syllable *ra*, but does not carry the acccompanying vowel *u*; it retains the inherent vowel *a*.

The ANGKA may also serve as a vowel carrier, as shown below. The boxed text shows ᝎᝊ, which is the syllable ᝎ MA followed by an ᝊ ANGKA carrying the ◌ VOWEL SIGN I.
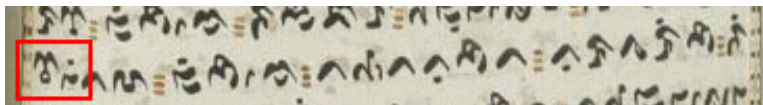
This text is to be read as ᝎᝎ *mami*. In this case, the two syllables have identical consonants, but only the second has a vowel sign.

The usage of ANGKA is based upon the practice of using the digit '2' as a mark of repetition. The form of ANGKA is derived from ꧏ U+A9CF JAVANESE PANGRANGKEP, which is itself based upon ٢ U+0662 ARABIC-INDIC DIGIT TWO. A similar system of syllable reduplication is used in Buginese. However, a separate ANGKA-type character has not been encoded for Buginese and the Unicode standard states that the Javanese PANGRANGKEP is to be used. As pairs of base letters and combining vowel signs belonging to different script blocks may complicate rendering, syllable identification, collation, and other processing, it may not be practical to use Javanese PANGRANKEP as a base letter in Makasar script contexts. For this reason, the ANGKA is proposed for encoding as a separate character in the Makasar block.

4.5 Digits

Script-specific digits for Makasar are not attested. The available manuscript sources show usage of two distinct sets of digits. The first set bears a strong resemblance to Latin digits and the second to Arabic-Indic analogues. Digits are used frequently and both sets occur concurrently in the sources. The two sets are shown below:

---

[1] Unless otherwise stated, all excerpts are from KIT 668-216 (see figure 3).

	Latin-like	Latin	Arabic-like	Arabic
zero		0		.
one		1		١
two		2		٢
three		3		٣
four		4		٤, ۴
five		5		٥, ۵
six		6		٦, ۶
seven		7		٧
eight		8		٨
nine		9		٩

There are particular digits within each set that differ from corresponding normative Latin and Arabic-Indic forms. Among the Latin-like digits, the 'one', 'seven', and 'nine' are distinct in that they are often written with a downward hook at the bottom right of the descender. This feature may be related to the use of a hook in some Balinese digits, eg. ᭓ U+1B53 BALINESE DIGIT THREE.

Among the Arabic-like forms, the 'zero' and 'five' are distinct from regular digits. The Arabic 'zero' resembles the Latin digit '0' more than the typical ٠ U+0660 ARABIC-INDIC DIGIT ZERO. It may be confused with ٥ U+0665 ARABIC-INDIC DIGIT FIVE. The Arabic-Indic form of 'five' does not resemble either ٥ U+0665 ARABIC-INDIC DIGIT FIVE or ۵ U+06F5 EXTENDED ARABIC-INDIC DIGIT FIVE; it is a alternate form that is found in Jawi sources.

Examples of usage of numbers are given below. The following excerpt shows the numbers 29, 250000, 30 written using Latin-like digits:

8
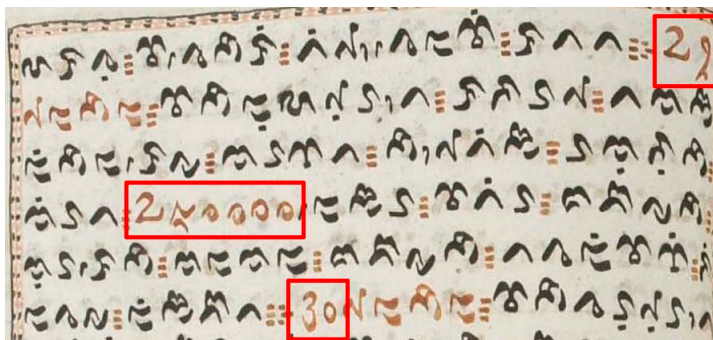
The following excerpt shows Latin-like digits in the numbers 19, 16, 67, 1670, and 17 (boxed in red), and Arabic-like digits in 15 and 1080 (boxed in blue):
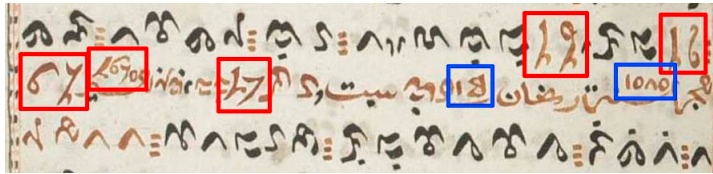
꞉ no we bu ru 19 ’e ra [30] 16
67 1670 hijîr [sic] pi bi re ru 17 ’a lo sabt bu la 15 ramaDân sanah 1080 hijr
pa ka na na ꞉[2]

The zero used in the number 1080 could be confused with ٥ U+0665 ARABIC-INDIC DIGIT FIVE, leading to the incorrect interpretation of ‘1080’ as ‘1585’. The correct value is derived from the usage of the number within the context of a Hijri date.

The numbers 1670, 15, and 1080 deserve further notice. They are written above what appear to be date and number signs:
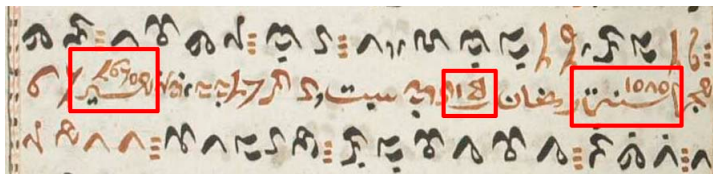
The number 1670 represents the Gregorian year 1670 and is written above a word that resembles the Arabic sequence هير or هسير, which may be a date mark for the Gregorian era. The number 15 is written above a line that might be the ؀ U+0600 ARABIC NUMBER SIGN. The number 1080 is written above the Arabic word سنة *sanah* (or a dotted form of ؁ U+0601 ARABIC SIGN SANAH) and represents the Hijri year 1080.

The representation of dates in Makasar documents is additionally notable as it suggests that usage of the two sets of digits is determined by linguistic context. The Arabic-like digits are restricted to Arabic-language environments and particularly for expressing dates of the Hijra era. The Latin-like digits are used for general purposes, but are used within Arabic-language contexts for writing non-Hijra dates, specifically those of the Gregorian calendar (see below for details).

Further research is needed for determining how to treat digits found in Makasar manuscripts. Forms such as the Latin-like ‘one’ and ‘nine’ may be distinctive enough to warrant separate encoding in the Makasar block, and the Arabic-like ‘zero’ and ‘five’ could be added as characters to the Arabic block. However, for the present, the Latin-like forms should be unified with Latin digits 0 .. 9. The Arabic-like forms should also be unified with ٠ U+0660 ARABIC-INDIC DIGIT ZERO .. ٩ U+0660 ARABIC-INDIC DIGIT NINE. The latter set should be specified as script extensions for Makassar. Space is reserved in the code block for adding script-specific digits, if needed.

The approach to handling usage of non-Arabic-Indic digits with ؀ U+0600 ARABIC NUMBER SIGN and ؁ U+0601 ARABIC SIGN SANAH as attested in the available sources needs to be better understood, but is out of scope for the present proposal.

---

[2] Transliteration courtesy of Christopher Miller.

4.6   Punctuation

Two punctuation signs are proposed for encoding:

⋮   MAKASAR PASSIMBANG

∴⁙   MAKASAR END OF SECTION

The Makassar PASSIMBANG is used for delimiting short segments of texts, or what may be considered 'sentences'. The mark consists of three dots oriented in a vertical column. It is similar to ⁙ U+1A1E BUGINESE PALLAWA.
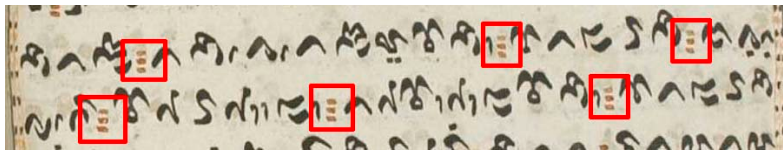
Longer segments of texts are denoted using the END OF SECTION. This mark consists of six dots oriented in the shape of a right-pointing triangle:
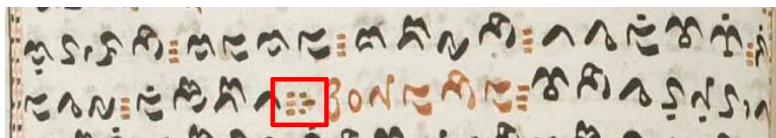
In some cases, the dots in the END OF SECTION mark are oriented in the form of a right triangle as ⁛. (TM Or545.232, reproduced in Jukes 2014). This form is a glyphic variant.
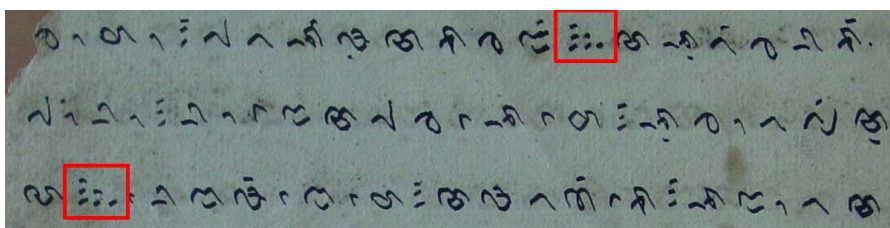
The excerpt below shows ∴⁙ END OF SECTION followed by a series of spaced dashes:
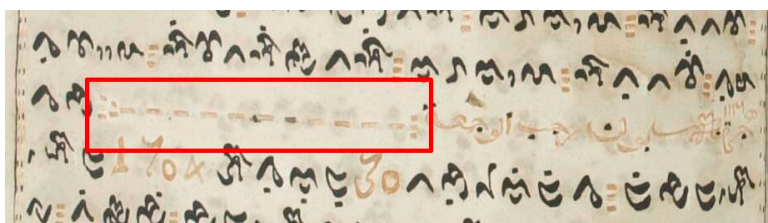
There is no need to encode a script-specific dash as the common – U+2012 FIGURE DASH, – U+2013 EN DASH, or — U+2014 EM DASH may be used in such cases.

10

4.7   Other sectioning marks

Other marks of punctuation used for indicating the end of section of text are attested in the available sources. These are described below, but are not proposed for encoding at the present.

A stylized representation of the Arabic word تمّت *tammat* 'it is complete' is used for indicating the end of a major section of text:
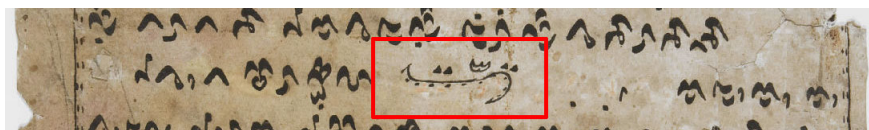
It is also written with decoration and flourish, as shown below:
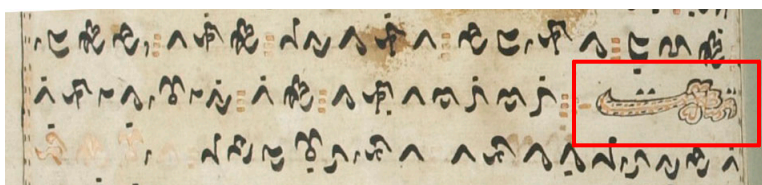
Here, *tammat* follows the ⁛ END OF SECTION mark:
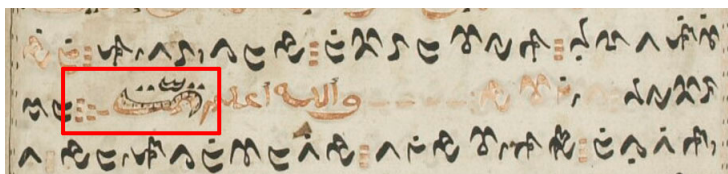
While it may be practical to treat تمّت as an atomic character — in order to preserve its function as a mark of punctuation with appropriate character properties — the word should be represented instead using the following sequence of Arabic letters:

ت U+062A ARABIC LETTER TA, م U+0645 ARABIC LETTER MEEM, ّ U+0651 ARABIC SHADDA, ت U+062A ARABIC LETTER TA

The ّ U+0651 ARABIC SHADDA may be removed from the sequence when representing instances of تمّت where the sign is not used: تمت.
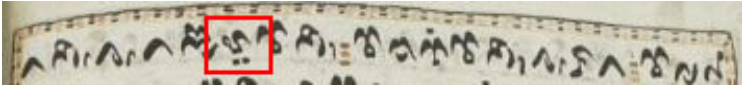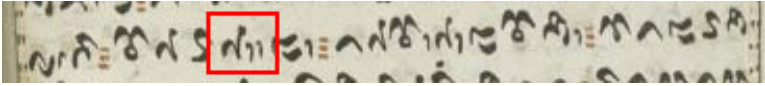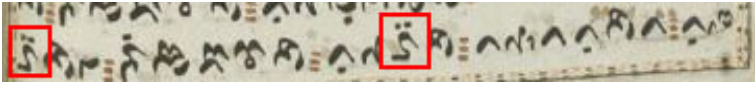
Another text-level punctuation mark is attested in a manuscript (microfilm at Australian National University) from the period 1834–1858 that is written is a variant form of the Makasar script (Jukes 2014: 5). It uses motifs resembling palm trees for marking the end of section:
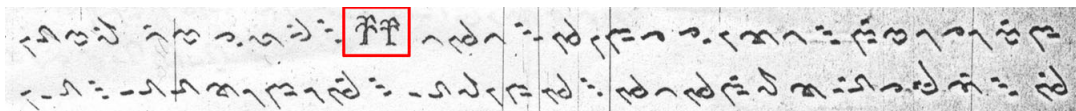
The tree motif is used only in this particular manuscript and there is no need to encode it as a separate character for Makasar. The existing character 🌴 U+1F334 PALM TREE from the 'Miscellaneous Symbols and Pictographs' block may be used.

4.8　Syllable abbreviation

Two contiguous and identical graphical syllables may be abbreviated by deleting the consonant of the second syllable and grouping its vowel sign with the first syllable, resulting in two vowel signs attached to a single base consonant. For example:

The abbreviated syllables shown above would be represented in encoded text as follows:

du·u　　Ꚉ　　<Ꚉ DA, ◌ VOWEL SIGN U, ◌ VOWEL SIGN U>

po·o　　ᨔ11　　<ᨔ PA, ◌1 VOWEL SIGN O, ◌1 VOWEL SIGN O>

li·i　　Ꚋ̈　　<Ꚋ LA, ◌̇ VOWEL SIGN I, ◌̇ VOWEL SIGN I>

4.9　Handling multiple vowel signs

In order to accommodate the system of syllable abbreviation described above, rendering engines should consider the contiguous occurrence of two of the same vowel sign as valid input. Moreover, the engine should provide appropriate spacing for sequences involving two prepending vowel signs:

Encoded sequence	Render output
kake　<ꠥ KA, ꠥ KA, ꠤ◌ VOWEL SIGN E>	ꠥꠤꠥ
kake·e　<ꠥ KA, ꠥ KA, ꠤ◌ VOWEL SIGN E, ꠤ◌ VOWEL SIGN E>	ꠥꠤꠤꠥ

If more than two vowel signs occur contiguously in an encoded sequence, then the additional signs would not be rendered appropriately with the base and would be displayed at their logical position in the encoded sequence with a dotted circle:

<ꠥ KA, ◌̇ VOWEL SIGN I, ◌̇ VOWEL SIGN I, ◌̇ VOWEL SIGN I>　　　　→　　ꠥ̈◌̇

<ꠥ KA, ꠤ◌ VOWEL SIGN E, ꠤ◌ VOWEL SIGN E, ꠤ◌ VOWEL SIGN E>　　→　　ꠤꠤꠥꠤ◌

Although the available sources do not show evidence of syllable abbreviation occuring with dissimilar vowel signs, sequences of such signs should be considered valid:

       ku·i   ꦨ   ‹ꦨ KA, ◌ VOWEL SIGN U, ◌̇ VOWEL SIGN I›

       ko·e   ꦨ   ‹ꦨ KA, ◌ VOWEL SIGN O, ◌ VOWEL SIGN E›

4.10   Linebreaking

Linebreaking generally occurs after an orthographic syllable; however there is the potential that syllables containing ◌ VOWEL SIGN E may be split across lines, such that the vowel sign remains the last character on the line and the consonant is written at the beginning of the next line. It is not clear at this time if such occurrences should be considered normative or idiosyncratic, and if there is an expectation for handling such occurrences. Hyphens or other marks indicating continuance are not used.

4.11   Collation

Collation for Makasar follows the sort order for Buginese:

       ꦨ KA < ꦨ GA < ꦨ NGA < ꦨ PA < ꦨ BA < ꦨ MA < ꦨ TA < ꦨ DA < ꦨ NA <

       ꦨ CA < ꦨ JA < ꦨ NYA < ꦨ YA < ꦨ RA < ꦨ LA < ꦨ VA < ꦨ SA < ꦨ A <

       ◌̇ VOWEL SIGN I < ◌ VOWEL SIGN U < ◌ VOWEL SIGN E < ◌̇ VOWEL SIGN O

The sort order for ꦨ ANGKA needs to be determined. If possible, the ANGKA should be sorted using the same weight as for the consonant letter of the preceding orthographic syllable. In cases where two identical consonants occur alongside a sequence of the same consonant and ANGKA, then the sequence containing the ANGKA should be sorted after the sequence containing the two identical consonants. A sample is given below:

       ꦨꦨ *kaka*, ꦨꦨ *kaka*, ꦨꦨ *kaki*, ꦨꦨ *kaki*, ꦨꦨ *kika*, ꦨꦨ *kika*, ꦨꦨ *kiki*, ꦨꦨ *kiki*, ꦨꦨ *kuka*,

       ꦨꦨ *kuka*, ꦨꦨ *kuku*, ꦨꦨ *kuku*, ꦨꦨꦨ *keke*, ꦨꦨꦨ *keke*, ꦨꦨꦨ *koko*, ꦨꦨꦨ *koko*

5   Character Data

5.1   Character Properties

Properties in the format of `UnicodeData.txt`:

```
11EE0;MAKASAR LETTER KA;Lo;0;L;;;;;N;;;;;
11EE1;MAKASAR LETTER GA;Lo;0;L;;;;;N;;;;;
11EE2;MAKASAR LETTER NGA;Lo;0;L;;;;;N;;;;;
11EE3;MAKASAR LETTER PA;Lo;0;L;;;;;N;;;;;
11EE4;MAKASAR LETTER BA;Lo;0;L;;;;;N;;;;;
11EE5;MAKASAR LETTER MA;Lo;0;L;;;;;N;;;;;
11EE6;MAKASAR LETTER TA;Lo;0;L;;;;;N;;;;;
11EE7;MAKASAR LETTER DA;Lo;0;L;;;;;N;;;;;
11EE8;MAKASAR LETTER NA;Lo;0;L;;;;;N;;;;;
11EE9;MAKASAR LETTER CA;Lo;0;L;;;;;N;;;;;
```

```
11EEA;MAKASAR LETTER JA;Lo;0;L;;;;;;N;;;;;
11EEB;MAKASAR LETTER NYA;Lo;0;L;;;;;;N;;;;;
11EEC;MAKASAR LETTER YA;Lo;0;L;;;;;;N;;;;;
11EED;MAKASAR LETTER RA;Lo;0;L;;;;;;N;;;;;
11EEE;MAKASAR LETTER LA;Lo;0;L;;;;;;N;;;;;
11EEF;MAKASAR LETTER VA;Lo;0;L;;;;;;N;;;;;
11EF0;MAKASAR LETTER SA;Lo;0;L;;;;;;N;;;;;
11EF1;MAKASAR LETTER A;Lo;0;L;;;;;;N;;;;;
11EF2;MAKASAR ANGKA;Lo;0;L;;;;;;N;;;;;
11EF3;MAKASAR VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;;
11EF4;MAKASAR VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11EF5;MAKASAR VOWEL SIGN E;Mc;0;L;;;;;;N;;;;;
11EF6;MAKASAR VOWEL SIGN O;Mc;0;L;;;;;;N;;;;;
11EF7;MAKASAR PASSIMBANG;Po;0;L;;;;;;N;;;;;
11EF8;MAKASAR END OF SECTION;Po;0;L;;;;;;N;;;;;
```

5.2   Linebreaking

Linebreaking properties in the format of `LineBreak.txt`:

```
11EE0..11EF1;AL   # Lo  [18] MAKASAR LETTER KA .. LETTER A
11EF2;AL          # Lo       MAKASAR ANGKA
11EF3..11EF6;CM   # Mn   [4] MAKASAR VOWEL SIGN I .. VOWEL SIGN O
11EF7..11EF8;AL   # Po   [2] MAKASAR PASSIMBANG .. END OF SECTION
```

5.3   Syllabic Categories

Syllabic categories given in the format of `IndicSyllabicCategory.txt`:

```
# Indic_Syllabic_Category=Vowel_Dependent
11EF3..11EF4   ; Vowel_Dependent      # Mn   [2] MAKASAR VOWEL SIGN I..VOWEL SIGN U
11EF5..11EF6   ; Vowel_Dependent      # Mc   [2] MAKASAR VOWEL SIGN E..VOWEL SIGN O

# Indic_Syllabic_Category=Consonant
11EE0..11EF0   ; Consonant            # Lo  [17] MAKASAR LETTER KA..LETTER SA

# Indic_Syllabic_Category=Vowel_Independent
11EF1          ; Vowel_Independent    # Lo       MAKASAR LETTER A

# Indic_Syllabic_Category=Consonant_Placeholder
11EF2          ; Consonant_Placeholder # Lo      MAKASAR ANGKA
```

5.4   Positional Categories

Positioning data for combining signs in the format of `IndicPositionalCategory.txt`:

```
# Indic_Positional_Category=Right
11EF6          ; Right                # Mc       MAKASAR VOWEL SIGN O

# Indic_Matra_Category=Left
11EF5          ; Left                 # Mc       MAKASAR VOWEL SIGN E

# Indic_Matra_Category=Top
11EF3          ; Top                  # Mn       MAKASAR VOWEL SIGN I

# Indic_Matra_Category=Bottom
11EF4          ; Bottom               # Mn       MAKASAR VOWEL SIGN U
```

5.5   Confusables

```
11EE4 MAKASAR LETTER BA      ;   1A0E BUGINESE LETTER NYA
11EE8 MAKASAR LETTER NA      ;   1A08 BUGINESE LETTER TA
11EF2 MAKASAR ANGKA         ;    A9CF JAVANESE PANGRANGKEP
11EF3 MAKASAR VOWEL SIGN I  ;   1A17 BUGINESE VOWEL SIGN I
11EF4 MAKASAR VOWEL SIGN U  ;   1A18 BUGINESE VOWEL SIGN U
11EF5 MAKASAR VOWEL SIGN E  ;   1A19 BUGINESE VOWEL SIGN E
11EF6 MAKASAR VOWEL SIGN O  ;   1A1A BUGINESE VOWEL SIGN O
11EF7 MAKASAR PASSIMBANG    ;   1A1E BUGINESE PALLAWA
```

6   References

Anderson, Deborah, et al. 2015. "Recommendations to UTC #144 July 2015 on Script Proposals" (L2/15-204). `http://www.unicode.org/L2/L2015/15204-script-recs.pdf`

Cummings, William P. 2007. *A Chain of Kings: The Makassarese Chronicles of Gowa and Talloq*. Leiden: KITLV Press.

Everson, Michael. 2003. "Final proposal for encoding the Buginese script in the UCS" (L2/03-191). `http://www.unicode.org/L2/L2003/03191-n2588-buginese.pdf`

Faulmann, Carl. 1880. *Das Buch der Schrift: Enthaltend die Schriftzeichen und Alphabete aller Zeiten und aller Völker der Erdkreises*. Zweite Vermehrte und verbesserte Auflage. Wein: Der Kaiserlich-Königlichen Hof- und Staatsdruckerei.

Holle, K. F. 1882. *Tabel van Oud- en Nieuw- Indische Alphabetten*. Bijdrage tot de palaeographie van Nederlandsch-Indië. Batavia: W. Bruining & Co.; 's Hage: M. Nijhoff.

Jukes, Anthony. 2014. "Writing and reading Makassarese". Presented at the International Workshop on Endangered Scripts of Island Southeast Asia, Tokyo University of Foreign Studies, February – March 2014. `http://lingdy.aacore.jp/doc/endangered-scripts-issea/anthony_jukes_paper.pdf`

Matthes, B. F. 1858. *Makassaarsche spraakkunst*. Amsterdam: Het Nederlands Bijbelgenootschap.

Miller, Christopher. 2010. "Unicode Technical Note #35: Indonesian and Philippine Scripts and Extensions". `http://www.unicode.org/notes/tn35/`

Pandey, Anshuman. 2015a. "Preliminary proposal to encode the Makassarese Bird Script" (L2/15-100). `http://www.unicode.org/L2/L2015/15100-makassarese-bird-script.pdf`

———. 2015b. "Proposal to Encode the Old Makassarese Script in Unicode" (L2/15-179). `http://www.unicode.org/L2/L2015/15179-old-makassarese.pdf`

Raffles, Thomas S. 1817. "The History of Java", vol. 2. London: Black, Parbury, and Allen.

Rahman, Nurhayati. 2014. "Sejarah dan dinamika perkembangan huruf lontaraq di Sulawesi selatan". Presented at the International Workshop on Endangered Scripts of Island Southeast Asia, Tokyo University of Foreign Studies, February – March 2014.
`http://lingdy.aacore.jp/doc/endangered-scripts-issea/nurhayati_rahma_paper.pdf`

7   Acknowledgments

	11EE	11EF
0	ꓰ 11EE0	ꓹ 11EF0
1	ꓩ 11EE1	ꓹ 11EF1
2	ꓨ 11EE2	ꓲ 11EF2
3	ꓪ 11EE3	◌̇ 11EF3
4	ꓳ 11EE4	◌̣ 11EF4
5	ꓴ 11EE5	ꓰ◌ 11EF5
6	ꓶ 11EE6	◌ꓱ 11EF6
7	ꓷ 11EE7	⦂ 11EF7
8	ꓸ 11EE8	⁙ 11EF8
9	ꓹ 11EE9	
A	ꓺ 11EEA	
B	ꓻ 11EEB	
C	ꓼ 11EEC	
D	ꓽ 11EED	
E	꓾ 11EEE	
F	꓿ 11EEF	

This script is known indigenously as Ukiri' Jangang-jangang and in English as Makassarese Bird Script.

Consonants

11EE0 ꓰ MAKASAR LETTER KA
11EE1 ꓩ MAKASAR LETTER GA
11EE2 ꓨ MAKASAR LETTER NGA
11EE3 ꓪ MAKASAR LETTER PA
11EE4 ꓳ MAKASAR LETTER BA
11EE5 ꓴ MAKASAR LETTER MA
11EE6 ꓶ MAKASAR LETTER TA
11EE7 ꓷ MAKASAR LETTER DA
11EE8 ꓸ MAKASAR LETTER NA
11EE9 ꓹ MAKASAR LETTER CA
11EEA ꓺ MAKASAR LETTER JA
11EEB ꓻ MAKASAR LETTER NYA
11EEC ꓼ MAKASAR LETTER YA
11EED ꓽ MAKASAR LETTER RA
11EEE ꓾ MAKASAR LETTER LA
11EEF ꓿ MAKASAR LETTER VA
11EF0 ꓹ MAKASAR LETTER SA
11EF1 ꓹ MAKASAR LETTER A

Consonant reduplicator

11EF2 ꓲ MAKASAR ANGKA

Vowel signs

11EF3 ◌̇ MAKASAR VOWEL SIGN I
11EF4 ◌̣ MAKASAR VOWEL SIGN U
11EF5 ꓰ◌ MAKASAR VOWEL SIGN E
11EF6 ◌ꓱ MAKASAR VOWEL SIGN O

Punctuation

11EF7 ⦂ MAKASAR PASSIMBANG
11EF8 ⁙ MAKASAR END OF SECTION

	Makassar	Buginese
KA	ꦨ	〃
GA	꧁	꧑
NGA	꧂	꧒
NGKA	—	꧓
PA	꧔	꧕
BA	꧖	꧗
MA	꧘	꧙
MPA	—	꧚
TA	꧛	꧜
DA	꧝	꧞
NA	꧟	ꧠ
NRA	—	ꧡ
CA	ꧢ	ꧣ
JA	ꧤ	ꧥ
NYA	ꧦ	ꧧ
NYCA	—	ꧨ
YA	ꧩ	ꧪ
RA	ꧫ	ꧬ
LA	ꧭ	ꧮ
VA	ꧯ	꧰
SA	꧱	꧲
A	꧳	꧴
HA	—	꧵

Table 1: Comparison of Makasar and Buginese consonants.

	Makasar	Buginese
VOWEL SIGN I	◌̇	◌̇
VOWEL SIGN U	◌̣	◌̣
VOWEL SIGN E	ꭲ◌	◌
VOWEL SIGN O	◌ꭲ	◌
VOWEL SIGN AE	—	◌

Table 2: Comparison of Makasar and Buginese vowel signs.

Makasar	Buginese
⦙	⸪
PASSIMBANG	PALLAWA
⁙	⧓
END OF SECTION	END OF SECTION
ꭲ	(٢)
ANGKA	(U+A9CF JAVANESE PANGRANGKEP)

Table 3: Comparison of Makasar and Buginese punctuation and other characters.

1662, 1670/1080 hijrah Chronicles of Gowa and Talloq

1729

1116 hijrah

1704/1141 hijrah

1739/1152 hijrah

p. 157 (?) above 'TAMMAT iyamite'

p. 157 (?) below 'TAMMAT iyamite'

p. 152 first to second *tammat*

Raffles (1817) *History of Java* "Ugi or Mengkásar Alphabet"

p. 152 above first *tammat*

p. 152 after third *tammat*

p. 152 second to third *tammat*

p. 153

p. 154

Campbell Macknight's microfilmed 19th century text (from Jukes Makassarese thesis)

Figure 2: Makasar character inventories from various sources. Compiled by Christopher Miller.

Figure 3: Excerpt from a hand-written book in the Makasar script (KIT 668-216). Image from Wiki-Media Commons, provided by the Tropenmuseum of the Royal Tropical Institute (KIT). Source: `http://commons.wikimedia.org/wiki/File:COLLECTIE_TROPENMUSEUM_Gedeelte_van_het_` `dagboek_van_de_Vorsten_van_Gowa_in_oud_Makassaarschrift_TMnr_668-216.jpg`.
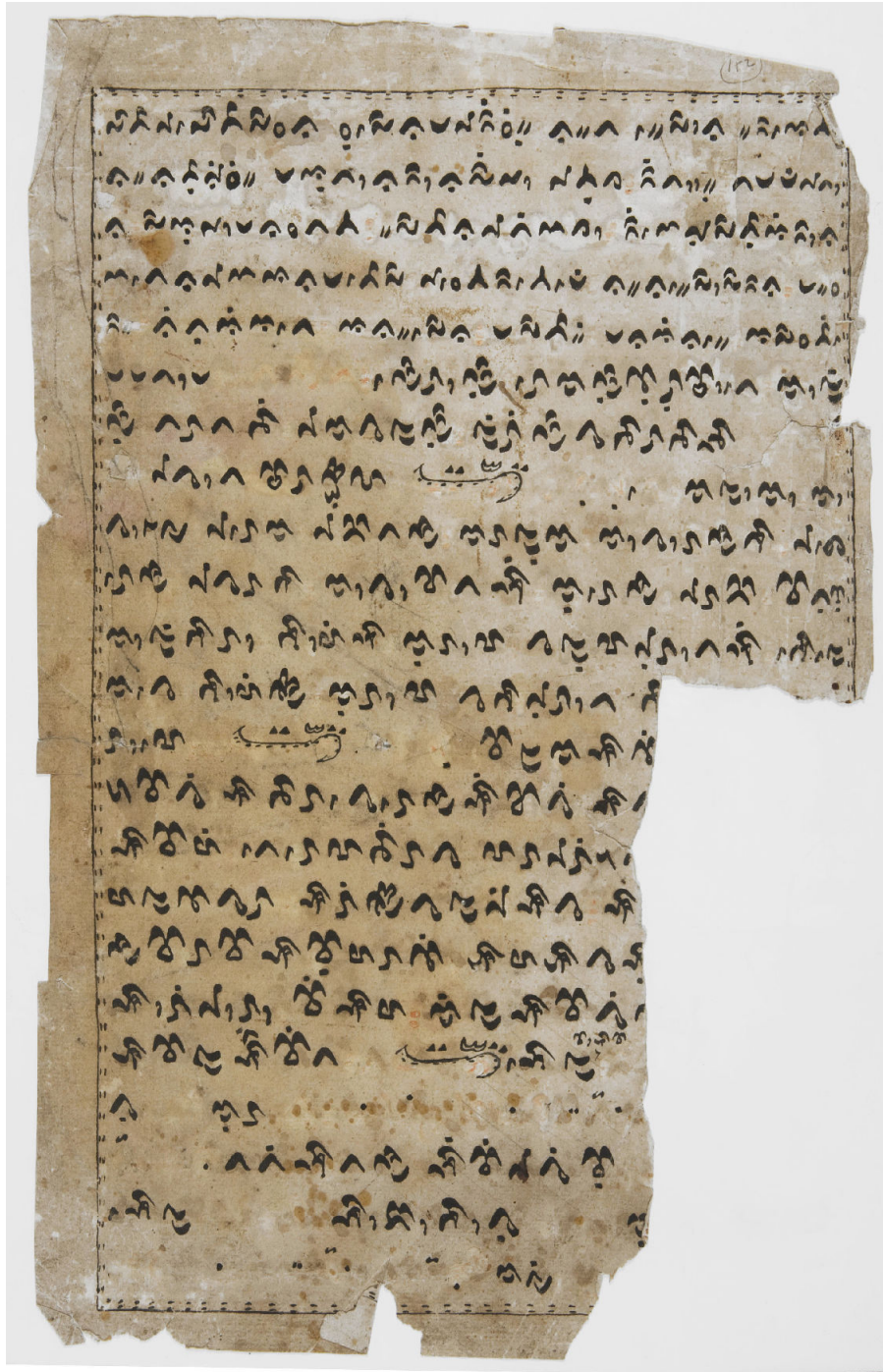
Figure 4: A folio containing text written in both the Buginese (first five lines and beginning of line six) and Makasar scripts (Tropenmuseum 668-216 no. 119). Image courtesy of Christopher Miller.

§ 37. De Maleische اڠك (aṅgka) (٢), zijnde niets anders dan het Arab. cijfer 2, dat men achter de woorden schrijft ten teeken, dat zij bij het uitspreken moe-

ten verdubbeld worden, wordt ook soms in het Makassaarsch overgenomen; bijv. ○◌◌ (en in het oude schrift ◌◌◌) = ○◌○◌ (en in het oude schrift ◌◌◌◌). Voorts heeft de Makassaar nog een ander middel om de verdubbeling eener enkele lettergreep aan te duiden, te weten: door den klinker tweemaal te schrijven; bijv.: ◌◌◌, in stede van ◌◌◌◌ (lôlo) *jong*; ◌◌◌, in stede van ◌◌◌◌ (tête) *brug*, enz. Met andere klinkers geschiedt zulks zelden, als ligtelijk aanleiding tot verwarring gevende.

Figure 5: Description of the ◌ MAKASAR LETTER ANGKA along with words printed in the Makasar ("oude schrift") and Buginese scripts (stitched together from Matthes 1858: 11, 12).

23
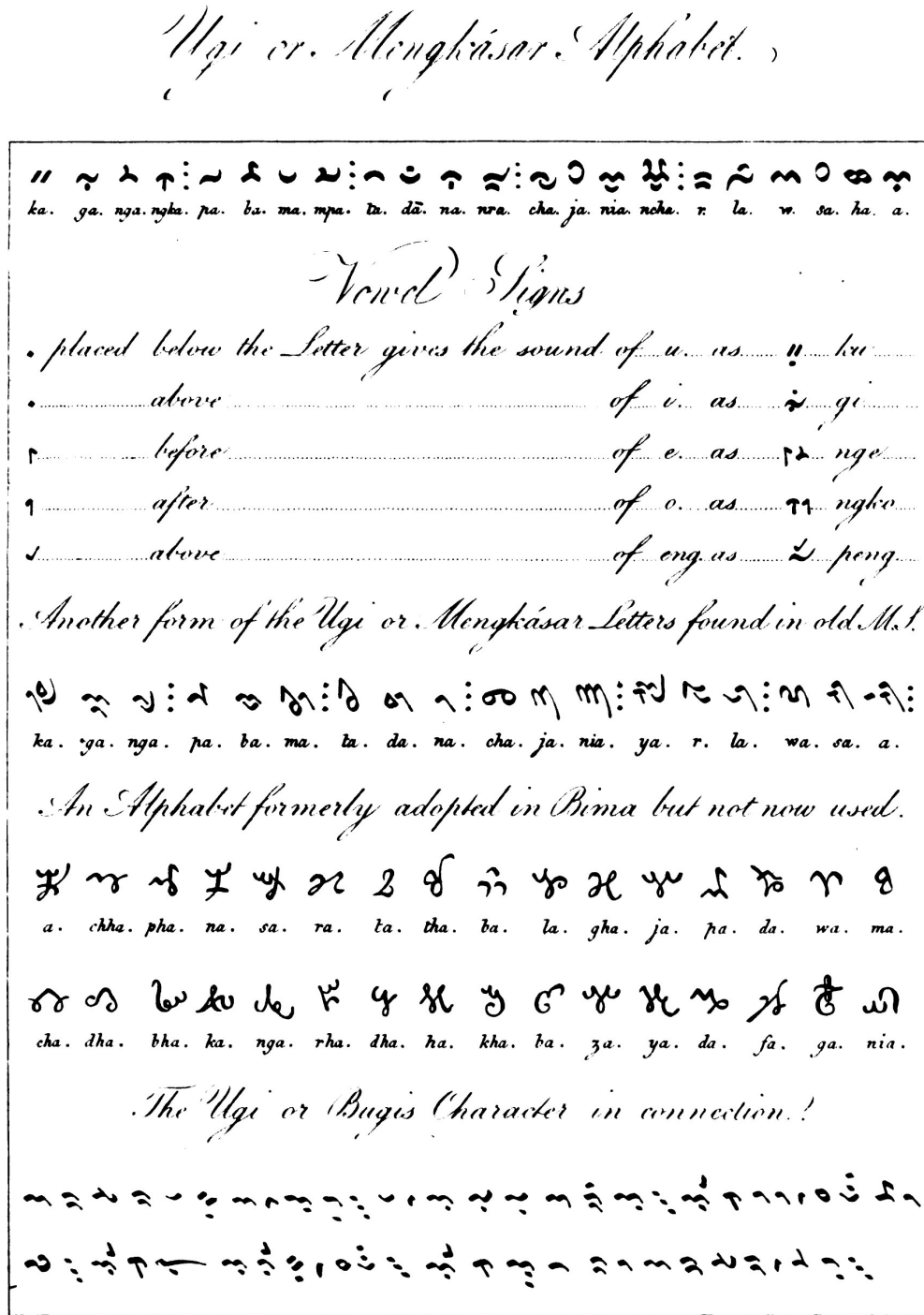
Figure 6: Chart showing scripts used in Makasar (from Raffles 1817, plate after p. clxxxviii) The Makasar script is shown under the heading "Another form of the Ugi or Mengkásar Letters found in old M. S.". The character repertoire shown here is identical to the proposed repertoire. Some glyph appear to be different, but the underlying graphical structure is evident. The Buginese script is shown at the top without a heading and at the bottom under the heading "Ugi or Bugis Character").

SCHRIFTEN AUF CELÉBES UND SUMATRA.

Maṅkāsar	Bugi		Battak		Redžaṅ	Lampuṅ	Wert
	Alt	Neu	Alt	Neu			
							a
							ka
							ga
							ṅa
							tša, tʼa
							dža, dʼa
							ńa
							ta
							da
							na
							pa
							ba
							ma
							ya
							ra
							la
							wa
							sa
							ha

Die Vokale der *Bugi* sind: ˙ *titi* oder *ana irate i*, . *titi* oder *ana irawa u*, ˻ *ana ri-boko ē*, ˻ *ana ri-yolo o* z. B.: ki, ku, kē, ko. Die *Battak*-Vokale sind: die Initiale a, i, u, in der Mitte und am Ende ○ *i*, ˃ *u*, ✗ *o*, ‿ *e*, ⁻ ˋ, *a* ist den Konsonanten inhärent, ‒ *e* steht am Anfang der Konsonanten (*bere*), ˘ am Ende oder über Vokalen (⁼ *pã*, —ō *pī*), ˃ wird auch verbunden, z. B. *bupusutu*.

158

Figure 7: Chart showing the Makasar ("Maṅkāsar") and related scripts (from Faulmann 1880: 179). Faulmann erroneously equates 𑻪 MAKASAR LETTER A with ∞ U+1A16 BUGINESE LETTER HA.

Volgorde der Letters	SUMATRA]							F. [CELEBES]				G. [BIMA]			H. BOR. NEO	J. PHILIP. PIJNEN
	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142
	Pasèmah	Redjang, Ommelanden Bengkoelen	Krocing	Man-deling	Ang-kola	Toba	Daèri	Makasser	Boegis	Oud Makass. Alph. volgens Matthes	Raffles	Oud in onbruik geraakt Alph., volgens Friedrich		Raffles	Jav. Alpht ꞏbalken kraton, Marta-poera	

Figure 8: Chart showing scripts from "Celebes" or Sulawesi (from Holle 1882: 11) Columns 136 and 137 show the Makasar script. The column showing transliteration ("Volgorde der Letters") has been stitched from the previous page in Holle.

Figure 9: Chart showing scripts from "Celebes" or Sulawesi (from Holle 1882: 20). Columns 136 and 137 show the Makasar script.

Volgorde der Letters	ATRA]					F. [CELEBES]				G. [BIMA]			H. BORNEO	J. PHILIPPIJNEN
	129	130	131	132	133	134	135	136	137	738	139	140	141	142
	jang Kroeïng	Man-deling	Angkola	Toba	Daёr	Makasser	Boegis	Oud Makass. Alph. volgens Matthes	Raffles	Oud in onbruik geraakt Alph. volgens Friedrich		Raffles	Jav. Alph. op balken kraton, Martapoera	
Dj Nj														
Rĕ														
Lĕ														
A														
_a														
E														
e_														
ai_														
I														
_i														
_î														
O														
_o														
Oe														
_oe														
_oê														
Ĕ														
_ĕr														
Klinkerdooder (Patĕn)														

Figure 10: Chart showing scripts from "Celebes" or Sulawesi (from Holle 1882: 29). Columns 136 and 137 show the Makasar script. The column showing transliteration ("Volgorde der Letters") has been stitched from the previous page in Holle.

Value	Bird script	South Sumatran	Javanese			Buginese
			Early 17th century	Modern. Balinese	Modern Javanese	
k						
g						
ng						
c						
j						
ny						
t						
d						
n						
p						
b						
m						
s						
l						
r						
y						
w						
ʔ						

Figure 11: Chart showing Makasar and related scripts (from Miller 2011: 44).

29

Figure 12: The left chart shows "Aksara Lontara Toa jangang-jangang" = "Old Lontara Bird Script", which is the Makasar script described here. The center chart shows "Aksara Lontara Baru" = "New Lontara Script' or Buginese. The right chart shows "Aksara Lontara Bilang-bilang" or the "Counting script". From a display at Balla Lompoa Museum, Sungguminasa, Gowa. Image from WikiMedia Commons, provided by Sandjaja Kosasih (User:Sanko). Source: http://commons.wikimedia.org/wiki/File:Lontara_script.jpg.

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for**
guidelines and details before filling this form.
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest** *Roadmaps*.

A. Administrative

1. **Title:** *Proposal to encode the Makasar script in Unicode*
2. Requester's name: *Anshuman Pandey / Script Encoding Initiative*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2 November 2015*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *Yes*
      Proposed name of script: *Makasar*
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block:
2. Number of characters in proposal: *25*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ____ B.1-Specialized (small collection) ____ B.2-Specialized (large collection) ____
   C-Major extinct ____ D-Attested extinct ____ E-Minor extinct *X*
   F-Archaic Hieroglyphic or Ideographic ____ G-Obscure or questionable usage symbols ____
4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document?
   b. Are the character shapes attached in a legible form suitable for review?
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the
      standard?
      Anshuman Pandey
      b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      Anshuman Pandey
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
      of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*

8. Additional Information:
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script
that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.
Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour
information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default
Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization
related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also
see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports
for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *No*
   If YES explain

2. Has contact been made to members of the user community (for example: National Body,
   user groups of the script or characters, other experts, etc.)? *Yes*
   If YES, with whom? *Ian Caldwell, Anthony Jukes, Sirto Koolhof, Christopher Miller*
   If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included? *No*
   Reference: *Historical script, currently not used*

4. The context of use for the proposed characters (type of use; common or rare) *Common*
   Reference:

5. Are the proposed characters in current use by the user community? *No*
   If YES, where? Reference:

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? *N/A*
   If YES, is a rationale provided?
   If YES, reference:

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *Yes*

8. Can any of the proposed characters be considered a presentation form of an existing
   character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to, or could be confused with, an existing character? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:

11. Does the proposal include use of combining characters and/or use of composite sequences? *Yes*
    If YES, is a rationale for such use provided? *Yes*
    If YES, reference: *See text of proposal*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:

12. Does the proposal contain characters with any special properties such as
    control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: